1954

Information theory and entropy in communication and scientific measurement

Hathway, Donald L.; Emerson, Duane L.

Monterey, California: U.S. Naval Postgraduate School

http://hdl.handle.net/10945/14262

INFORMATION THEORY AND ENTROPY IN COMMUNICATION AND SCIENTIFIC MEASUREMENT

DONALD L. HATHWAY
AND
DUANE L. EMERSON

INFORMATION THEORY AND ENTROPY IN COMMUNICATION AND SCIENTIFIC MEASUREMENT

* * * * * * * * *

Donald L. Hathway
and
Duane L. Emerson

INFORMATION THEORY AND ENTROPY
IN COMMUNICATION AND
SCIENTIFIC MEASUREMENT

by

Donald L. Hathway
Lieutenant, United States Navy

and

Duane L. Emerson
Captain, United States Army

Submitted in partial fulfillment
of the requirements
for the degree of
MASTER OF SCIENCE
IN
PHYSICS

United States Naval Postgraduate School
Monterey, California

1 9 5 4

This work is accepted as fulfilling

the thesis requirements for the degree of

MASTER OF SCIENCE

IN

PHYSICS

from the

United States Naval Postgraduate School

PREFACE

In December 1953 during the closing lectures in a first course in Statistical Mechanics conducted by Dr. M. S. Watanabe of the Department of Physics, U. S. Naval Postgraduate School, a brief acquaintance was made with "Information" in connection with entropy. The possibility of relating, in a more definite fashion, a rudimentary appreciation of the fundamental significance of entropy with another study of extensive application -- the transfer or recovery of "intelligence-bearing" symbols or signals -- was intriguing. Early in this year, fortified with the expression A log B and the encouragement of Dr. Watanabe the authors set forth into the realms of the rapidly-developing Information Theory. This paper presents a few of the landmarks and boundaries encountered in this broad field where the underlying unity is sometimes obscured by the diversity of application.

The authors wish to express their appreciation to Dr. M. S. Watanabe and to Dr. Randolph Church for their patient assistance and contagious enthusiasm.

TABLE OF CONTENTS

INTRODUCTION

Summary

The entropy concept is discussed with reference to statistical mechanics and thermodynamics. After a demonstration by examples of the fundamental principles of the Communication branch of Information Theory, "information" and "entropy" are compared as to mathematical form and as to fundamental relationship. Various viewpoints on scientific measurement are set forth to suggest the similarity between a measurement system and a communication system. A theory of scientific information is briefly considered, and the features of a measurement system are referred to related aspects of a communication system. Entropy is discussed again as it pertains to measurement; it is seen that the necessity for measurement prevents Maxwell's demon from violating the second principle for the model assumed. The violation would require the procurement of "free" information which itself would entail a violation of the second principle.

CHAPTER I

ENTROPY

1. Introduction

In considering the behavior of physical systems, it is important to be able to maintain a balance sheet of the various energy transformations which occur in natural processes. This accounting, however, gives little indication as to the type and extent of energy conversions which may be realized in practice. Such limitations are given expression in the second principle of thermodynamics; they are not inherent in the first principle. From empirical and mathematical-model viewpoints, a measure of the tendency to proceed exhibited by a physical system when it is free to change, has been formulated in the <u>entropy</u> concept.

In addition to its engineering significance, entropy is an important concept in modern communication theory and in Scientific Information Theory. Since appreciation of the interplay of the engineer or the scientific observer with the system under investigation is enhanced by an understanding of the nature of entropy, developments of the latter concept will now be considered. Essential to this discussion of the entropy function are terms delimited as follows:

(a) A body whose properties have specified values is said to be in a certain <u>state</u>, and the variables which are chosen to specify the properties are called <u>parameters of state</u>.

(b) The term *system*, as used in thermodynamics, refers to a definite quantity of matter bounded by some closed surface.

(c) A system can exchange energy with its surroundings by the performance of mechanical work or by a "flow of heat." If conditions are such that no energy interchange can take place, the system is said to be *isolated*.

(d) When an isolated system is left to itself and the parameters of state are measured at various points throughout the system, it is observed that although these quantities may initially change with time, the rates of change become smaller and smaller until eventually no further observable (observable with the instruments and scale of measurement employed) change occurs. This final steady state of an isolated system is called a state of *thermodynamic equilibrium*.

(e) A *process* is any event in a thermodynamic system in which a redistribution or transformation of energy occurs and is evidenced by a change in the thermodynamic coordinates of the system.

(f) A *reversible* process is one that may be described by a succession of equilibrium states, or states that depart only infinitesimally from equilibrium. In order for a process to be reversible, it is essential that it be possible to return the immediate system and any others associated with it from their last to their initial state

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

in exactly inverse order, and that it be possible to return from final to original form, location, and amount all the energy which was transformed during the process.

(g) An irreversible process is one that does not meet the specification for reversibility. On the thermodynamic scale all known natural processes are irreversible. The full requirement of irreversibility is, that it is impossible, even with the assistance of all agents in nature, to restore the exact initial state everywhere in the system once the process has taken place. The definition of irreversibility implied above in no manner demands that this phenomena extend to all scales of investigation. This point will be considered at length subsequently. However, it is interesting to note here, that friction, which is an important contributor to irreversibility, is not required in the systematic description and explanation of phenomena on the astronomical or the atomic scales of investigation.

2. Thermodynamics and Entropy

According to Planck, (39) the only clear way of showing the significance of the second principle is to base it on facts by formulating propositions which may be proved or disproved by experiment. Listed below are a few of those propositions:

1. It is in no way possible to completely reverse any process in which heat has been produced by friction.
2. It is in no way possible to completely reverse any process in which a gas expands without performing

work or absorbing heat.

3. If there is heat conduction between two bodies at different temperature, it is in no way possible to convey this heat back without leaving any change whatsoever.

4. It is in no way possible to reverse the process of diffusion. (Essentially the same as 2)

Upon the introduction of the term "reverse" in the above propositions, we are met with the concept of irreversibility. The full requirement of irreversibility is, that it is impossible, even with the assistance of all agents in nature, to restore everywhere the exact initial state when the process once takes place. Upon the above propositions rest the whole structure of the second law of thermodynamics. If any one of them could be found to be actually reversible within the confines of the afore-stated definition, then, because of their interrelation, all of them would be capable of being reversed. Since they all represent actual observable processes in nature, then were they reversible, the second principle would be untrue.

The next step in consideration of the second principle is the realization that it furnishes a relation between the quantities connected with initial and final stages of any natural cyclic process. In reversible cyclic change, the initial and final states are identical; whereas in irreversible cyclic processes, there is some difference between states as pointed out by the second principle. Then from the mathematical viewpoint, the distinction between initial and final

states consists of an inequality.

With this thought in mind, we turn to the mathematical inequality developed by Clausius on an empirical basis:

$$\oint \frac{d'Q}{T} \leq 0. \qquad (1.0)$$

Applying this relation to a cyclic process, all portions of which were considered reversible, Clausius arrived at an expression for entropy change,

$$\int_1^2 dS = \int_1^2 \frac{d'Q}{T} = S_2 - S_1. \qquad (1.1)$$

Through further employment of this relation, this time to a cycle, part of which is irreversible, it may be determined that the <u>change</u> in entropy for an isolated system left to itself is always positive. This determination is accomplished as follows:

> Consider an isolated thermodynamic system in equilibrium in state 1. As a result of a natural (and hence irreversible) process, the system moves from equilibrium state 1 to equilibrium state 2. By means of a reversible process, the system is then returned to state 1. Taken together, the two processes constitute a cycle which as a whole is irreversible.
>
> From the Clausius inequality,
>
> $$\oint \frac{d'Q}{T} < 0,$$
>
> or writing the integral as the sum of two integrals,

[illegible]

$$\int_1^2 \frac{d'Q}{T} + \int_2^1 \frac{d'Q}{T} < 0. \tag{1.11}$$

Since the system was isolated during the change from state 1 to state 2, no heat could enter or leave the system. Hence,

$$\int_1^2 \frac{d'Q}{T} = 0. \tag{1.12}$$

However, in order to return to state 1 and complete the cycle, the exchange of heat and work with elements outside the system must take place. Since this is a reversible process,

$$\int_2^1 \frac{d'Q}{T} = S_1 - S_2. \tag{1.13}$$

From inequality 1.11,

$$S_1 - S_2 < 0 \quad \text{or} \quad S_2 - S_1 > 0.$$

3. Statistical Mechanics and Entropy

Before applying probability procedures, we must first discover how well thermodynamic systems lend themselves to an approach of this type. There are important properties of matter which can not be derived from gross thermodynamic considerations alone. We can go beyond these limitations only by making hypotheses regarding the nature of matter, and by far the most fruitful of such hypotheses is that matter is composed of discrete particles. For simplicity, the discussion that follows will be limited to an ideal monatomic gas, specifically to a finite volume containing a large number of independently acting mass points in continual motion. Based on the proposed system at hand, we are immediately aware of the limitations

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

of the observer who cannot deal with individual units of the system, but rather, only with measurable data such as density, volume, and temperature. He will be referred to as the macro-observer. Let us hypothecate a super observer who can see every molecule and relate their individual positions in space and velocity; he then is the micro-observer. The latter has the mechanical idea of state, the former the statistical average idea of state. The correlation, then, between the mechanical and statistical approaches to thermodynamic systems stems from the fact that a <u>given</u> macro idea of state be characterized by many different "mechanical" ideas of state. Because the macro-observer has only measurable data on which to base his calculations, because this measurable data depends upon the particular macro-state of the system, and finally due to the fact that any given macro-state can be characterized by many possible micro-states, we here find adequate basis for the usefulness of probability theory as the means of description of the given system. It must be understood that due to the chaotic movement of the molecules of the gaseous system, all <u>a priori</u> possible micro-states are not realized in nature. [Klein 30]

If different portions of the system were at varying macro-states, then the system would be described as being in molar order. If, however, the entire system has the same macro-state, then we consider the system as being in a state of molar disorder. We can consider molar disorder as being synonomous with settled, and molar order as being synonomous with unsettled. A simple analogy is that of a swimming pool being filled at one end with water much colder than that in the pool.

That end originally will be cooler than other portions, hence the entire pool might be considered as being in an unsettled or more ordered state. Given time, with no interference from the outside, all portions of the pool will reach the same temperature. At this stage the pool is in a settled or less ordered state. Here, by Nature's own process, we have a transformation from order to disorder, unsettled to settled, and reach thermal equilibrium throughout. It is found that the number of micro-states is smaller for the unsettled than the settled state, thus indicating a trend toward a greater number of micro-states. Considering each micro-state as a complexion, we can define the probability W of a state as the number of complexions in that state.

A more specific description of this natural tendency to attain a more probable state may be achieved by a consideration of the "H" function of statistical mechanics. Boltzmann's H-theorem which demonstrates the actual tendency for the molecules of a system to approach their equilibrium or most probable state employs a function [Tolman 49]

$$H = \sum_i n_i \log_e n_i + \text{Constant}, \tag{1.2}$$

where n_i is the number of molecules in the different cells in coordinate momenta space.

This expression may be written as

$$H = -\log_e P + \text{Constant}, \tag{1.3}$$

where $\log_e P$, as derived from Maxwell Boltzmann's Statistics, may be expressed by

$$\log_e P = n \log_e n - \sum_i n_i \log_e n_i + C. \qquad (1.4)$$

Boltzmann's H-Theorem states that H decrease algebraically toward its minimum possible value as the system approaches the condition of equilibrium.

A generalized form of the H-theorem was developed by Gibbs in which an ensemble of systems is considered rather than a single system with which Boltzman's initial H-theorem was concerned. The generalized approach by Gibbs, a more powerful method than that of Boltzman, defines a similar quantity $\bar{\bar{H}}$ which also decreases with time. The quantum mechanical analogue of $\bar{\bar{H}}$ may be considered in variational fashion to yield the following expression

$$-\delta \bar{\bar{H}} = \frac{\delta \bar{\bar{E}}}{\Theta} + \frac{1}{\Theta}\left(\bar{\bar{A}}_1 \delta a_1 + \bar{\bar{A}}_2 \delta a_2 + \cdots\right) \qquad (1.5)$$

in which $\bar{\bar{E}}$ is the mean energy, $\bar{\bar{A}}_n$ denotes the mean values of the external forces calculated over the members of the ensemble, δa_n denotes the variations made in the external coordinates, and Θ is a distribution parameter. The above equation is then compared to a derived form of the combined first and second principles:

$$\delta S = \frac{\delta E}{T} + \frac{1}{T}\left(A_1 \delta a_1 + A_2 \delta a_2 + \cdots\right) \qquad (1.6)$$

The similarity of these two forms makes it reasonable to correlate the thermodynamical quantities S and T with $-\bar{\bar{H}}$ and Θ as follows:

$$S = -k\bar{\bar{H}} \qquad\qquad T = \frac{\Theta}{k} \qquad (1.7)$$

(and this particularly in view of the similarity in the tendency for S to increase in natural processes as previously discussed). Thus we see that the quantity S may be expressed as

$$S = -k \sum_n P_n \log_e P_n \tag{1.8}$$

where P_n equals the (exact) probabilities for the true energy states n in the canonical ensemble which we take as representing the equilibrium, and k is a constant with the dimensions of energy over temperature which turns out to be Boltzmann's constant or the perfect gas constant per molecule. When we consider the special case of a system regarded as being with equal probability in one or another of a group of W micro-states between which no distinction is made on the basis of macro-scopic measurements, this relation reduces to

$$S = k \log_e W. \tag{1.9}$$

4. Comparison

Thus we have indicated the development of the concept of entropy from two standpoints which have been shown to be compatible with the behavior of physical systems. From the statistical view, the entropy is expressed as a function proportional to the probability of a state of a particle system: S equals k $\log_e$ W. It is of interest to note that the thermodynamic expression for entropy $dS \geq \int \frac{dQ}{T}$, based on the Clausius inequality, assumes the character of a pure number if the temperature is measured in thermodynamic work units and hence is compatible with the notion of probability. Both expressions for entropy define a function, which, as a parameter of state, does not decrease for any natural process in an isolated system. However, the

statistical approach more aptly explains the behavior of the system. Irreversibility, for example, is not inherent in the dynamical motions of the individual particles but in their combined mean effect. It is to be noted that the concept of entropy does not appear in the considerations of basic kinetic theory since this is based on the dynamical treatment of the motion of individual particles within the limitations of the assumptions of the kinetic model of matter.

Further comparison of the two approaches to the mathematical development of entropy discloses additional points in which they differ. For instance, thermodynamic entropy of a system is empirically defined for equilibrium states only, whereas from the statistical standpoint, the entropy of a system can be determined for any state whether or not equilibrium has been attained. Also, only changes in entropy for reversible processes can be computed with equation (1.1) while statistically the entropy can be determined for the initial and final states of any process, reversible or irreversible, the difference being the change. Finally, a comparison of basic equations shows the empirical derivation to be a differential whose integral gives the change in entropy from the initial to the final state or the entropy referred to any arbitrary standard; the statistical entropy concept makes possible the calculation of absolute entropy.

The method of Gibbs and of Boltzman have been mentioned previously in connection with entropy as defined by statistical mechanics. An additional method, that of Darwin and Fowler, provides an interesting check on the other ways of looking at the problem and on certain questionable approximations, e.g., the wide use of Sterling's formula

for N! can be avoided. This latter method approaches the problem through the use of mean values. The assumption is, that in a very long period of time, a system would pass through all accessible states, the time spent in each state being proportional to the number of complexions of that state. [Lindsay 31]

Having investigated the development of the entropy concept, we are now in a position to summarize the more important features of entropy and the second principle of thermodynamics.

a) There exists in nature a quantity which changes always in the same sense in all natural processes. [Planck 39]

b) The impossibility of an uncompensated decrease in entropy seems to be reduced to an improbability. [Klein 30] by Gibbs.

c) Net growth of entropy in all bodies participating in an occurrence means that the system as a whole has experienced an irreversible change of state. This change is of course in harmony with the first law of energy but this growth gives additional information as it indicates the <u>direction</u> in which a natural process occurs. [Klein 30]

d) When all the participating bodies of the system are considered, every natural event is marked by an increase in the number of complexions of the system. This is the most precise physical statement of the second law and covers the whole domain of science. [Klein 30] by Planck.

e) Entropy is a measure of the <u>range in phase</u> of the system. Greater entropy goes with a greater ranging of the molecules over molecule space, ... A non-equilibrium state is then one in which full use is not being made by the system of the phase-space ranges

that are open to it under the conditions to which it is subject so that its behavior exhibits less phase range than in the state of equilibrium. [Kennard 29]

f) A recent article by Muses suggests another interpretation of entropy. A state of 100% entropy represents a condition of complete lack of disturbance of the electromagnetic and gravitation medium. The increase of entropy attendent to natural processes might then be attributed to the elastic hysteresis loss of an elastic medium.

Certainly further study of Muses' article is required before describing entropy in these terms. It is known, however, that elastic hysteresis effects depend upon previous states as well as upon the instantaneous conditions; and that the hysteresis loss is related to the rate of loading and unloading, being less at slow rates. This time dependence might correspond to the approach to reversibility in thermodynamics when a process is conducted at ever slower rates.

g) Finally we give a mathematical concept which covers the whole domain of physics: "Any function whose time variation always has the same sign until a certain state is reached and is then zero may be called an entropy function." [Klein 30]

One final observation is in order. We have shown, using the laws of mechanics and certain hypothesis, that statistical mechanics is able to define a quantity whose mathematical behavior is the same as that of the entropy of thermodynamics. The latter says that ΔS is equal to or greater than zero; the former says ΔS is equal to or greater than zero with overwhelming probability. There is a possibility of <u>almost</u> <u>accounting</u> for the second principle by mechanical

[illegible]

[illegible]

[illegible]

reasoning. Thus one might be willing to extrapolate this partial success and to state that ultimately thermodynamic entropy and its statistical mechanical analogue will be found to be identical. Although statistical mechanics theoretically provides a finite though small possibility of reversing the second principle, in view of the model assumed and assumptions made in the statistical development, and because Planck's rigorous propositions supporting the second principle have not been invalidated in practice, it is inadvisable to accept the identity proposed above.

CHAPTER II

INFORMATION THEORY - COMMUNICATION

1. Introduction and Definitions

Modern communication (or information) theory is the confluence of two branches of science. One branch starts with the earliest attempts of mathematicians, such as Kelvin and Heaviside, who applied quantitative descriptions to problems of signal transmission. The second started in the twenties of this century with the first theories of noise and broadened into the statistical theory of communication when Wiener, Kolmogoroff, and Shannon conceived not only the noise, but also the messages as part of statistical series. Thus "pure" communication theory appears as the application of two branches of mathematics to communication processes -- analysis on the one hand, probability theory on the other -- and forms itself a new branch of applied mathematics. As such, it requires a solid foundation of physical laws and empirical data whenever it is applied to any practical problem. General notions of the problems concerned with efficient message formulation, transmission, and reception have been known for some time. It is the great achievement of Shannon that he was able to replace the rather vague meaning of the word "information" by a more precise definition which allows the assignment of a numerical value to an amount of information and hence makes possible the mathematical analysis of the content of messages and of a wide variety of situations which may be considered similar in nature.

The procedure used to develop the theory mathematically attacks the problem from the communication standpoint since there the theory found its first application. As a consequence, many of the terms utilized are those peculiar to the engineering communication field. A listing of additional definitions appears in Appendix I; definitions essential to the following discussion are inserted here.

(a) fixed constraint - In telegraphy for example, four symbols: dot, dash, letter space, and word space are used. The organization of the code forbids a letter space or word space to follow a letter space or word space. Such restrictions are called fixed constraints.

(b) probability constraint - Languages have constraints controlled by usages. All letters in the basic alphabet do not occur with the same frequency. Furthermore, pairs of letters (digrams) and three letter systems (trigrams) have varying frequencies. This coupling process continues up through word-word combinations also having certain frequencies of occurrence. The use of a language to transmit information thus involves the consideration of probability constraints.

(c) ergodicity - the existence of a unique (i.e., independent of the initial condition), non-vanishing probability of each symbol or sequence of symbols appearing in infinitely long messages engendered by a set of intersymbol correlation probabilities. [Watanabe 51]

(d) noise - signals which are not coherent with any signals to which meaning is assigned in any transmission system.

(e) binary digit - a unit employed in the measurement of information which determines a single choice between equiprobable alternatives. The logarithmic base of two is conventional and convenient in practice.

(f) message - a particular selection from among the symbols or code elements constituting a code which has been made in conformity with the restrictions applicable to the occasion.

(g) information - in the most general sense, as that which adds to any structure, abstract or concrete, of which the features correspond in some sense with those of another structure.

(h) communication system - a system comprised of those elements which are essential for the initiation, transmission, and reception of intelligence-bearing signals.

Communication systems can be roughly classified into three categories; discrete, continuous, and mixed. The discrete system is one in which both the message and the signal are a sequence of discrete symbols. Continuous and mixed systems will not be discussed in this paper.

(i) signal element - a code element of a form which is suitable for transmission over the medium.

In order to analyze a communication problem by means of mathematical methods, a precise definition must be given which will allow a numerical value to be assigned to a sequence of intelligence-bearing symbols. Therefore, we shall employ the following definition to meet this requirement:

The amount of information received in a message is defined as

$$\text{Amount of Information Received} = \log_2 \frac{P_{ea}}{P_{eb}} \qquad (2.0)$$

where P_{ea} is the probability at the receiver of the event <u>after</u> the message is received, and P_{eb} is the probability at the receiver of the event <u>before</u> the message is received. The use of the logarithm makes the amount of information in independent messages additive. Equation (2.0) will now be applied to several examples to demonstrate its suitability.

2. Application of Equation

(a) <u>m events, m symbols</u>

Consider the problem of transmitting over a noiseless and discrete system the names of all residents of New York City and their ages. In the noiseless case, the numerator in equation (2.0) is unity. Assume that possible ages vary from one to one hundred inclusive. Let p_{12} be the probability that the age "twelve" will be sent. Then the amount of information received in a message wherein "twelve" was the transmission $= -\log_2 p_{12}$. Since all ages are assumed independent of one another, the total information can be found by adding the information reported by separate symbols. If

there are m different people, then m X p_{12} equals the number of transmissions of age "twelve" equals N_{12}. Then N_{12} X (- log p_{12}) equals total information reported for these N_{12} symbols. Summing over all possible ages gives the total information which is

$$-m\sum_{i=1}^{i=100} p_i \log_2 p_i \;;$$

then the average information per symbol is

$$-\sum_{i=1}^{i=100} p_i \log_2 p_i \,. \qquad (2.1)$$

All that is necessary for this equation to hold is that m be a large number.

(b) <u>An Ergodic Sequence</u>

Consider the problem of a long ergodic sequence consisting of m symbols, the symbols being taken from an alphabet of L symbols. [Goldman 23] Divide the sequence into r groups each consisting of q symbols, the number q being chosen large enough to surpass the inter-symbol influence. Thus $r = m/q$. Since the alphabet has L symbols, there will be L^q different groups q symbols in length. Let $s = L^q$. The different symbol groups are specified as group 1,2,..... s and N_1, N_2 ..... N_s are the number of each in the original sequence. Then $r = m/q = N_1$ plus N_2 plus N_3 ..... plus N_s. The total number of combinations of r things taken N_1, N_2, ... N_s at a time is $Mm = \dfrac{r!}{N_1!\, N_2! \,\ldots\ldots\, N_s!}$, Mm being the total number of different possible arrangements of the m symbols in the original sequence. Probability constraints enter the picture and thus for

[illegible]

the derivation to hold $N_1 = p_1 \times r$, $N_2 = p_2 \times r$ ..... $N_s = p_s \times r$. Taking the logarithm of M_m and using Sterling's approximation which is for large numbers

$$\ln N! = (N + ½) \ln N - N + ½ \ln 2\pi,$$

it is found that

$$\ln M_m = -r \sum_{i=1}^{s} p_i \ln p_i - \left(\frac{s-1}{2}\right) \ln r - ½ \sum_{i=1}^{s} \ln p_i - \left(\frac{s-1}{2}\right) \ln 2\pi.$$

Since m was chosen very large, and q has a small range, r will be large; hence, all but the first term may be dropped so that

$$\ln M_m = -r \sum_{i=1}^{s} p_i \ln p_i .$$

Choosing any one particular sequence, we find that the probability of that sequence is

$$P = p_1^{(p_1 r)} \, p_2^{(p_2 r)} \, \ldots\ldots \, p_s^{(p_s r)} = \frac{1}{M_m} .$$

Then

$$\ln P = r \sum_{i=1}^{s} p_i \ln p_i = - \ln M_m.$$

Since at this time we are dealing only with messages, let us modify equation 2.0 to read

$$\text{The Amount of Language Information Received} = K \ln \frac{\text{Probability at the receiver of the message after transmission received}}{\text{Probability at the receiver of the message before transmission received}} \qquad (2.15)$$

It can be seen that for a noiseless channel, the amount of language information received =

$$-K \ln P = K \ln M_m = -r K \sum_{i=1}^{s} p_i \ln p_i . \qquad (2.2)$$

Since a very long sequence was broken down into r groups, each of these groups can be considered as the basic unit and therefore,

$$\text{The Amount of Language Information Received per Unit} = -K \sum_{i=1}^{s} p_i \ln p_i . \qquad (2.3)$$

(c) <u>Telegram Problem - multiple form symbols</u>

The fundamental nature of this expression for information received will be emphasized by one final example. Brillouin (11) proposed a problem similar to the one just considered. A simplified model of a telegram consisting of only dots and blanks was chosen. If G positions were available, they would be filled with N_1 dots and N_2 blanks such that G equals N_1 plus N_2. Due to possible variations of pulses there might occur P_1 types of dots and P_2 types of blanks. Since the G positions would all be filled but only a maximum of <u>one</u> pulse or blank can fill a given position (cell), generalized Fermi-Dirac Statistics are applicable.

The total number of ways of filling the G cells is $\dfrac{G!}{N_1!\,N_2!}$ and recalling the various types of pulses, we must multiply this expression by $P_1^{N_1} P_2^{N_2}$. Thus there are

$P_1^{N_1} P_2^{N_2} \dfrac{G!}{N_1!\,N_2!}$ total complexions. Any one given message may be realized in $P_1^{N_1} P_2^{N_2}$ ways. Hence the probability of a specific message is

$$\frac{P_1^{N_1} P_2^{N_2}}{P_1^{N_1} P_2^{N_2} \Big/ \dfrac{G!}{N_1!\,N_2!}} \quad \text{or} \quad \frac{N_1!\,N_2!}{G!} = P.$$

Utilizing equation (2.2) above for a noiseless channel, we see that the amount of language information received per cell is $-K \ln P$.

[illegible]

By means of Sterling's approximation this can be shown to be

$$I = -K \sum_{i=1}^{2} p_i \ln p_i \quad \text{where } p = N/G.$$

If it is seen that the p_i here correspond to those in the previous example, that the G cells correspond to the r groups, and finally that P_1, P_2 represent types of pulses having the same significance whereas in the previous example all groups were distinct, then the parallelism between Brillouin's problem and the previous one is apparent. In passing, it is of interest to note that the analysis of the problem given by Brillouin involves the use of "physical entropy" and "message entropy" which will be discussed in a subsequent portion of this paper.

3. System with Noise

The three problems thus far considered have been limited to the noiseless case or system. The receiver is sure that he has received the exact message sent, and therefore the numerator of equation (2.0) becomes unity. Attention will now be given to the more usual and more involved case -- the system with noise. Here the probability of the message or event after receipt of the transmission is less than unity. To show the effect of noise in the system on the information received, equation (2.0) will be used with the following notation:

p_i = probability that i will be the transmitted message.

p'_j = probability that j will be the received message.

p_{ij} = probability that j will be the received message if i is the transmitted message.

From these notations it is seen that

[illegible]

[illegible]

[illegible]

$$\sum_i p_i = 1, \; \sum_j p'_j = 1, \; \sum_i p_i p_{ij} = p'_j, \; \sum_j \sum_i p_i p_{ij} = 1,$$

$p_i p_j = 1$ if i and j are the same,

$p_i p_j = 0$ if i and j are not the same, and

$P_A(i) = p_i$. These are alternate notations wherein the A refers to the transmitter.

Denoting the receiver as B, let $P_B(i)_j$ be the probability that i is transmitted and j received. Before receiving the message, we know the probability that i will be transmitted is p_i, and that the probability that i will be transmitted and j received is $p_i \times p_{ij}$. After the received message is found to be j, this factor is increased by $1/p'_j$ since all cases where j is not the received message are excluded. Then

$$P_B(i)_j = \frac{1}{p'_j} p_i p_{ij}.$$

Equation (2.0) now reads

$$\begin{array}{c}\text{Amount of Information} \\ \text{Received} \\ \text{Relative to } j\end{array} = \log_2 \frac{P_B(i)_j}{P_A(i)} = \log_2 \frac{\frac{p_i p_{ij}}{p'_j}}{p_i} .$$

[Goldman 23]

In order to visualize the significance of this equation, it will be applied to a simple problem. Suppose we have binary symbols (0) and (1). Let $P_A(1) = .5$ and $P_A(0) = .5$. During transmission, noise affects the system to the extent that 1/100 of the transmitted symbols are received incorrectly [transmitted (0) received as (1) and vice versa]. Referring to the above notation, $p_{11} = .99 = p_{00}$;

$p_{01} = .01 = p_{10}$.

Since $$\sum_i P_i P_{ij} = P'_j,$$

$$P'_1 = .5 \times .01 + .5 \times .99 = .5 \text{ and}$$

$$P'_0 = .5 \times .99 + .5 \times .01 = .5.$$

Then $$P_B(1)_1 = \frac{.5 \times .99}{.5} = .99,$$

$$P_B(0)_1 = \frac{.5 \times .01}{.5} = .01,$$

$$P_B(1)_0 = \frac{.5 \times .01}{.5} = .01,$$

$$P_B(0)_0 = \frac{.5 \times .99}{.5} = .99.$$

a) if a (1) is transmitted and a (1) received,

Amount of I. rec'd = $\log_2 \left[\frac{.99}{.5}\right]$ = .985 binary digits per symbol,

b) if a (1) is transmitted and a (0) received,

Amount of I. rec'd = $\log_2 \left[\frac{.01}{.5}\right]$ = -5.64 binary digits per symbol.

Were the system noiseless, we would have had instead of the above, $\log_2 1/.5 = 1$ binary digit per symbol. Thus as in case a), even though the transmitted symbol is the one received, the fact that because of noise, a (0) could have actually been transmitted, in effect reduces the amount of information received. Case b) presents an interesting example due to the negative results. [Woodward (55) termed this "deception"]. To the recipient, the probability of (1) being transmitted is initially .5. Upon receiving (0), the a posteriori probability of (1) having been transmitted is reduced to .01 despite the fact that it was the transmitted symbol. Thus the transmission in the presence of noise has made the state even less probable than it was to begin with.

3. Ambiguities in the Phrase "Amount of Information."

The results of the examples just considered appeared in the form "average amount of information per symbol (or unit)." This would seem to imply that a long message always reports a greater amount of information than a short message. However, a brief message carrying an account of a rare event may contain a greater amount of information than a long message dealing with a common occurrence. The above ambiguity arises from the fact that "amount of information" may be given different interpretations. Before attempting to resolve this ambiguity, the interpretations of this phrase (or of terms closely related to it) which have been assigned by writers in the field in Information Theory will be recounted.

> Signals are complexes of data transmitted from one physical system to another, and they convey information only if they are not predictable from the data previously received. Thus incomplete knowledge of the future, and also of the past of the transmitter from which the future might be constructed, is at the very basis of the concept of information. On the other hand, complete ignorance also precludes communication; a common language is required, that is to say an agreement between the transmitter and the receiver regarding the elements used in the communication process..... The information of a message could now be defined as the 'minimum number of binary decisions which enable the receiver to reconstruct the message, on the basis of the data already available to him.' These data comprise both the convention regarding the symbols and the language used, and the knowledge available at the moment when the message started.
>
> In this form, however, the definition is a counsel of perfection, of little practical use and even partly self-contradictory. It requires individual discussion of every given situation, which may not be exactly repeatable. In order to make it practical and meaningful, there must be added the important clause that the definition applies only to the average of a great number of samples, taken at random from a statistically homogeneous or 'ergodic' series. By this assumption (which is extremely difficult to define in a completely rigorous way) the previous contradiction is avoided.

On the other hand, it is clear that information in the exact sense of communication theory is far more restricted than the vague concept which goes by this name in everyday life. It may be also mentioned that this definition has nothing to do with the "value" of information. It is a measure of the minimum effort or cost by which the message can be transmitted, not of its importance or consequences. [Gabor 21]

..... the reader must be warned that there is some risk of confusion between three different quantities which are all likely to be measured in the same units. There is first the information capacity of a communication channel, which for telegraphic purposes could be measured in binary digits per second. Then there is the information content of a signal as transmitted, which for a telegraphic signal could be again measured in binary digits. Finally there is something which is proportional to the degree of confidence of the recipient of the message that he has received it correctly. [Bell 3]

Hartley purposely confined his attention to capacity, which is a quantity characteristic of a physical system. He was aware that "psychological factors" might have to be taken into account when defining an actual quantity of information, and assumed that these factors would be irrelevant to the communication engineer. The especially interesting feature of present day theory is the realization that information content differs from capacity not so much for psychological reasons as for purely statistical reasons which can very profitably be taken into mathematical account. Shannon's statistical treatment does indeed explain the "psychological" aspects of information to a quite remarkable degree..... When a communication is received, the state of knowledge of the recipient or "observer" is changed, and it is with the measurement of such changes that communication theory has to deal..... The information content of a message may be defined as the minimum capacity required for storage. [Woodward 55]

The effect of the information in a message is to change the probability concerning a situation, as far as the receiver of a message is concerned, from its value before the message is received to what is usually a larger value after the message is received. In a general way, it would appear that the amount of information in the message should be measured by the extent in the change in probability produced by the message.....languages, as we all know, are used in transmission channels to transmit information. The first step in this process is the coding of the messages at the information source into the (English) message alphabet. Thus, for example, an event occurs at the information source, and its description for transmission is its coded equivalent in the message alphabet. We have used the same word "message" for both the

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

event and its description. When it is desirable to make a distinction, we shall call the former, the event, and we shall call its coded equivalent in the language the message. [Goldman 23]

> In order that the message should carry information, there must be a probability at some receiver concerning the occurrence of the event which can be changed by the reception of the message..... According to our terminology, if p is the probability of a particular message in a language and $\bar{p}$ is the probability of the event which it describes, then we will say that (-log p) is the amount of language information in the message and (-log $\bar{p}$) is the amount of semantic information in the message. [Goldman 23]

It is apparent that there is a lack of a unified basis for discussion in the above points of view. Bell warns the reader of this in the first quotation. Woodward mentions "information content" and "information capacity" bringing out the belief that "phychological factors" enter into the measurement of quantity of information. In the next quotation, Goldman points out the manner in which "amount of information" alters the probability concerning a situation, and differentiates between an event and the message describing it. The event is distinguished from the message by calling the logarithm of its probability the amount of semantic information, while the logarithm of the probability of the message describing the event is termed amount of language information. In another quotation, this one being from Gabor, information is defined in terms of the number of binary decisions which the receiver must make to reconstruct the message.

Considerable unity and clarity are achieved in the basic analysis of the scope of information theory by the statements of Mackay (34):

> General information theory is concerned with the problem of measuring changes in knowledge. Its key is the fact that we can represent what we know by means of pictures, logical statements, symbolic models, or what you will. When we

receive information, it causes a change in the symbolic picture, or representation, which we would use to depict what we know.

We shall want to keep in mind this notion of a representation, which is a crucial one. Indeed, the subject matter of general information theory could be said to be the making of representations.....the different ways in which representation can be produced, and the numerics both of the production processes and of the representations themselves.

By throwing our spotlight on this representational activity, we find ourselves able to formulate definitions of the central notions of information theory which are operational, with more resultant advantages than that of current respectability. In any question or debate about "amount of information," we have simply to ask: "what representational activity are we talking about, and what numerical parameter is in question?" And we eliminate most of the ground for altercations.....or we should do so if we are careful enough!

We can cover, I think, all technical senses of the term "information" by defining it operationally as that which logically enables the receiver to make or add to a representation of that which is the case, or is believed or alleged to be the case..... Preconceived possibilities: that is the key phrase in communication theory. The communication engineer assumes that the receiver possesses a filing cabinet of prefabricated representations, so that for him a signal is an instruction to select one from the assembly or "ensemble" of possibilities already foreseen and provided for. His representational activity is not a constructional but a selective operation..... Amount of selective information is evidently a measure of the statistical rarity of a representation and has no direct logical connection with its form or content, except in cases where these affect its statistical status. One word which was unexpected could yield more selective information to a receiver than a whole paragraph which he knew he would receive.

Now it is evident that in any situation in which what is observed is thought of as specifying one out of an ensemble of preconceived possibilities, the amount of selective information so specified can in principle be computed. The concept has, therefore, a much wider domain of usefulness than that of communication theory. The point is that it is always a relevant parameter of a communication process, because successful communication depends on symbols having significance for the receiver, and hence on their being already in some sense prefabricated for him. The practical difficulty, of course, is to estimate the proportions of the appropriate ensemble, when these are determined by selectively -- and even unconsciously -- assessed probabilities. [Mackay 34]

The ambiguity arising with respect to the amount of information

reported by long and short messages is resolved if one considers the problem in the light of the interpretation given by Mackay. The notions of preconceived possibilities, representation, selection, and statistical rarity together point the way toward a reasonable explanation. An unusual event would stand very low on the observer's ladder of preconceived possibilities. Consequently, in his representation, he would assign the occurrence of that event a low probability. The opposite procedure would be applied to an event of common occurrence. Equation (2.0) contains the a priori probability P_{eb} which can be considered as that assigned in the observer's representation. Since P_{eb} appears in the denominator of the logarithmic term, the receipt of a message describing an event to which a small a priori probability was assigned would yield a greater amount of information than would the receipt of a message describing a less statistically rare event. Hence, the amount of information per symbol attained from a given message depends not only upon the number of symbols in the message, but also upon the statistical rarity of the event recounted by the message. With reference to the differentation between semantic and language information given by Goldman, the above discussion pertains only to events which are in the category of semantic information. However, reference can be made to equation (2.15), and the same discussion applied to language information if we speak of statistical rarity with regard to particular sequences or configurations of symbols or units.

At the very basis of the analysis of information theory quoted from Mackay are the concepts of <u>representation</u> and <u>selection</u>, the

[illegible]

latter being concerned with the choice, from the assembly of possibilities making up the representation, of one designated by an incoming signal. On the basis of this choice and the statistical rarity of that possibility chosen, the receiver derives an amount of selective information. Earlier in this chapter three example problems were explained through the use of equations (2.0) and (2.15) which involve, respectively, semantic and language information. They can also be discussed in the light of selective information. Example (a) dealt with the ages of the population of New York City. The receiver knows in advance the symbol significance and the message form of the communication system. He also knows that the incoming message will pertain to population age data for New York City. Based upon whatever knowledge he might have of age distribution for a normal population, the receiver forms a representation, assigning a priori probabilities to each age of the assembly of ages. Then each transmission of a name and age would instruct him to select that age from his representational assembly. Through the use of equation (2.0) the receiver can compute the amount of semantic information received. Since, however, he has used the a priori probabilities from his representation to compute this amount, the receiver has also determined the amount of selective information received. Example (b) (ergodic sequence) and example (c) (simplified telegram) are concerned with language information. Here again the receiver is familiar with the symbol significance and the message form of the communication system, but, because of lack of additional knowledge, he is unable to assign a probability of zero to "non-pertinent" symbol sequences. Hence, his prior representation

consists of the possible selections available to the message originator and their respective probabilities. Then the receipt of any particular sequence instructs the receiver which one to choose from his predetermined "ensemble" of possibilities. By substituting the corresponding a priori probability into equation (2.15) and carrying through the necessary computation, the receiver determines the amount of language information received. In this case, amount of selective information corresponds to amount of language information since the a priori probability used in the computation was that selected from the receiver's representation.

Thus, from its application to the three examples, its ability to explain the apparent ambiguity arising when considering the amount of information received from long and short messages, and, finally, the manner in which it is able to bring together the interpretations of information given by various authors in the field of information theory, the analysis of information theory proposed by Mackay appears to be the most extensive and fundamental.

[illegible]

[illegible]

CHAPTER III

INFORMATION THEORY AND ENTROPY

1. Introduction

Chapters I and II have considered the following formulae:

Entropy	Amount of Information
$S = -k \sum_n p_n \log_e p_n$	$I = -K \sum_i p_i \log_e p_i$
$S = k \log_e W$	$I = K \log_e M_m$

The question now arises that although I and S are expressed in a similar form, are the phenomena related? It has been stated in the consideration of entropy that the greater the entropy of a state the higher the probability of that state. With reference to an amount of information, it may be deduced from equation (2.0) that the less probable a certain event is, in the representation of the receiver, the more information a message carrying news of that event conveys. Thus

$$\text{Amount of Information} = \log_2 \frac{1}{\text{probability of an event before transmission is received}} = \log_2 \frac{1}{P_B}.$$

Let I take on an increment ΔI and P_B an increment ΔP_B;

$$\text{Then } I + \Delta I = \log_2 \frac{1}{P_B + \Delta P_B} \quad \text{or} \quad 2^{I + \Delta I} = \frac{1}{P_B + \Delta P_B}.$$

It follows that, if $\Delta P_B > 0$ then $\Delta I < 0$.

[illegible]

2. Negentropy Principle

> Bell (3) states -- Information is the negative of entropy. This means that the potential information content of any pattern can be assessed mathematically by the same process used to define the entropy..... Having progressed from entropy as a parameter of heat engines to entropy as a measure of disorder, there is no difficulty in taking the further step of relating a decrease of entropy to an increase of information.

If a book were set up in type, it would be in an ordered state and would provide a means of conveying information. Were this type broken up, the "entropy" of the system of letters would be greatly increased while the information would be destroyed. This example points to the relationship between information and the negative of "entropy" or negentropy.

All discussion thus far with regard to information and entropy brings to light one important peculiarity exhibited by the information formula. If the amount of information behaves like <u>negentropy</u>, why is its formula, ignoring the constants k^K, identical with that of entropy? As a stepping stone in the development of the reasoning behind this apparent ambiguity, let us first of all add to the list of definitions of entropy the following: <u>entropy is a measure of roughness of knowledge with the observer included in the system</u>.

> Woodward (55) states -- the information function is really a measure of prior ignorance in terms of prior probabilities. When the message state is known, probabilities become certainties, the ignorance is removed, and information correspondingly gained.

The observer is the recipient of the information, and it is upon his representation of the system that the information gain will have application.

Consider again equation (2.0). Suppose the probability of receiving a given message is p_x. At the same time, let the communication channel be subject to noise so that the probability at the receiver after the message is received is p_{1x}. Then we would have that the amount of information received = $\log \frac{p_{1x}}{p_x} = \log \frac{1}{p_x} - \log \frac{1}{p_{1x}}$.

Based on statements above, this may be written as the amount of information received = prior ignorance - final ignorance. Or in terms of the extended definition of entropy implied above, the amount of information received = initial entropy - final entropy. In the noiseless case, P_{1x} = 1 and the amount of information received becomes equal to the initial entropy.

(a) Mine Problem

To further illustrate this extended concept of entropy, let us apply it to this simple example. Suppose an armored battalion intelligence officer learns that a specified area to his front contains a powerful anti-tank mine. In referring to his map, he finds that the area under consideration extends over 32 co-ordinate squares as shown by the solid lines in the diagram below.

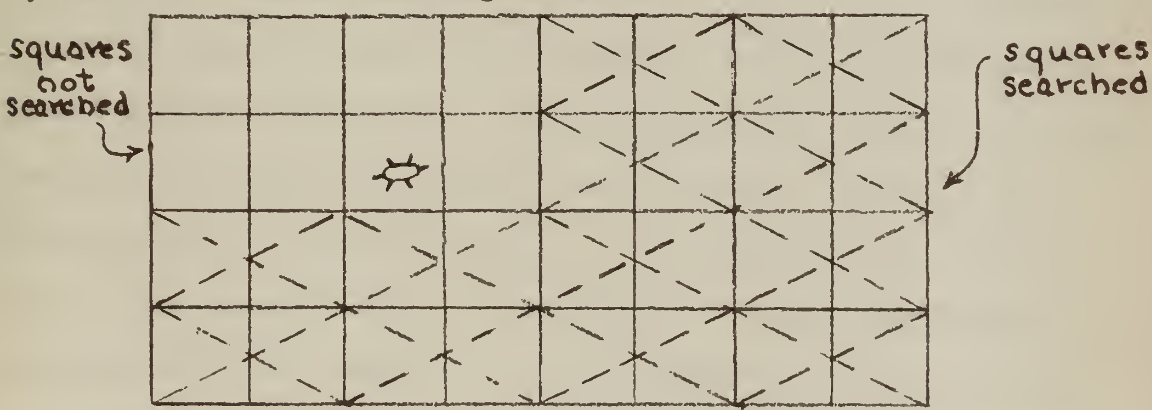

Fig. 1 Mine field in which one ground mine is known to be concealed

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]	[illegible]	[illegible]

[illegible]

[illegible]

The circumstances are such that he considers the mine has an equal chance of being in any one of the 32 squares. Thus the probability of the mine being in a given square is 1/32. Then his initial ignorance is $\log_2 32$ or 5 binary digits. A mine detection team is ordered into the area and returns with the report that the mine was not discovered. Their search covered the twenty-four squares indicated on the preceding diagram. The final ignorance of the intelligence officer is $\log_2 8$ or 3 binary digits. Therefore the information gain is 2 binary digits. From the standpoint of "entropy," with the initially specified area and the intelligence officer making up the system, it may be stated that the system in its initial state had 5 binary digits of "entropy" with respect to the mine location. Upon the receipt of information, the "entropy" of the system was <u>reduced</u> to three binary digits. In other words, the information gain resulted in an "entropy" loss; the information gain acted to produce a more ordered state of affairs -- i.e. 24 squares were opened for passage or occupancy.

(b) <u>Physical Entropy Example</u>

Another example, proposed by Brillouin (12), of the relationship between entropy and amount of information on a more involved scale should suffice to substantiate the proposition that information is the negative of entropy. The example was previously discussed from a different point of view in Part II. Let us suppose, in this case, that we are concerned with the probability distribution in phase space of electrons, originally in thermodynamic equilibrium, along a telegraphic wire. We assume that the passage of an assemblage of

electrical impulses of finite duration will affect only a small sub-ensemble of the total group. The choice of this sub-ensemble is determined by the constraints imposed on the overall system through the specification of a message of a given length which contains a certain number of impulses of different types. From the observer's end, since he originally knows of these constraints, the number of ways in which these electrons can be distributed after the passage of a particular assemblage is given by the number of possible distributions of the impulses. Referring to Problem (c) in Part II, this is:

$$P_1^{N_1} P_2^{N_2} G! / N_1! N_2! .$$

Since all of these messages are equally probable, the physical entropy of the sub-ensemble of electrons is given by:

$$S_{phys} = k \ln P_1^{N_1} P_2^{N_2} G! / N_1! N_2! . \qquad (3.1)$$

(by Stirling's formula) $$S_{phys}/_{per\,cell} = k\left(-P_1 \ln \frac{P_1}{\bar{P}_1} - P_2 \ln \frac{P_2}{\bar{P}_2}\right). \qquad (3.2)$$

However, since in the given conditions of the problem, P_1 represents pulses (dots) which vary in shape, intensity, and length, and P_2 has the same connotation with regard to the dashes, the transmission of any one of the above configurations can be received in $P_1^{N_1} P_2^{N_2}$ ways. Then, since the observer is unaware of which of the $P_1^{N_1} P_2^{N_2}$ ensembles was transmitted, these become representative of his uncertainty as to the final configuration of the electron sub-ensemble brought about by the instantaneous passage of one of these possible received groups of impulses. We can denote the final or message

[illegible]

$$[illegible]$$

[illegible]

$$[illegible]$$

$$[illegible]$$

[illegible]

entropy of the sub-ensemble by:

$$S_{m/\text{per cell}} = k\left(p_1 \ln P_1 + p_2 \ln P_2\right). \tag{3.3}$$

The difference between the physical and message entropies, as shown below:

$$S_{phys} - S_m = -k \sum_{i=1}^{i=2} p_i \ln p_i = I_g, \tag{3.4}$$

yields a measure of the information received concerning the distribution of the electrons making up the sub-ensemble. The gain of information reduces the observer's statistically characterized physical entropy of the system. Here we have obtained a physical measure for information only because our assembly (that being the sub-ensemble of electrons along a cable) was defined to be originally in thermodynamic equilibrium. Our results show that information corresponds to a negative term in the final entropy of the system;

$$S_m = S_{phys.} - I_g.$$

If the system were noiseless (i.e., only one kind of dot and dash thus making $P_1 = P_2 = 1$), S_m would be zero. This result implies absolute certainty on the part of the receiver. Then $S_{phys} = I_g$.

3. Unit Difficulties

Although in the simple examples discussed "information" and the negative of entropy have exhibited a similarity in form and a correspondence in behavior, there is still much confusion in the literature on information theory as to the relationship between the two quantities. The following comments are proposed in an attempt to clear up some of the ambiguity. As Bell (3) points out, there is room for argument as to whether there is a "real" or "physical" connection between the two.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

With formula (2.0), two as the convenient base for logarithms was arrived at by taking K = one in the more general equation:

$$I = K \log_n \frac{\text{Probability at the receiver of the event after message}}{\text{Probability at the receiver of the event before message}}.$$

This is convenient and customary in information theory as previously pointed out. When information is related to the entropy of a given thermodynamic system by writers such as Brillouin, K is set equal to Boltzmann's constant. To quote Bell (3):

> .....information is measured by a pure number, in general the product of a logarithm by a frequency or probability, whereas negentropy includes Boltzmann's constant and will only be a pure number if k is merely a numerical constant .....(without dimensions). For example, it is definite that kT represents an energy, but it has not been usual to apportion the dimensions of energy between k and T. If negentropy is identical with information, it is T alone which must be identified with energy, and k, measured in ergs per degree centigrade, is a pure number which has the value 1.37 X 10 exp -16 and is twice the factor needed to convert the scale of T from degrees centigrade to ergs per degree of freedom.....

Bell admits that in his previous discussions he has tacitly employed the point of view of regarding entropy as a mathematical abstraction representing pattern. He then concludes that there is a case for making entropy a mathematical abstraction (a pure number) rather than an energy function and that if this is admitted, the identity of information with the negative of entropy follows immediately.

It appears that the difficulty proposed by Bell stems from the idea that if physical entropy is made a dimensionless quantity, that it is no longer physical entropy -- that is, a characteristic of thermodynamic state indicating a definite "natural tendency." Here it appears, however, that the basic consideration is one of measurement

and units employed. If in the Clausius equation for entropy (which is empirically derived and hence the foundation for other derivations) the temperature is measured in <u>absolute work units</u>, then entropy becomes a pure number but one nevertheless related to the energy of a system. Wilson (53), in an excellent discussion on dimensions, has made the point as follows:

> Turning to thermal quantities, we may use as a substitute for temperature the Willard Gibbsian modulus which is equal to $k\,\Theta$, where k is Boltzmann's constant and Θ is the Kelvin work scale temperature. If we represent this temperature substitute by Θ' its dimensions are those of energy. Entropy would thus acquire the dimensions of a "pure number;" since its nature appears to be that of a probability, this would seem very appropriate.

Measurement of temperature in work units may be accomplished by using a Carnot engine as the thermometer along with some arbitrary assumptions as to standard and range. It will be recalled that when the generalized H-theorem of Gibbs was expressed in a form to give a parallel result with the thermodynamic entropy, the Boltzmann's constant made its appearance in the expression for the statistical analogue of entropy $S = -k\,\bar{\bar{H}}$. If temperature is measured in work units, then the Boltzmann's constant becomes a pure number, and hence both thermodynamic entropy and its statistical mechanical analogue become pure numbers which are nonetheless related to a system with given a priori constraints.

4. Conclusion

The confusion in identifying an amount of information with entropy stems from the extended use of the word entropy. Clarity of discussion may be achieved by calling the expression: $-K \sum_i p_i \log p_i$

an entropy of a probability distribution in which

$$\sum p_i = 1 \quad \text{and} \quad p_i \geq 0.$$

This is a general nomenclature which is not limited to a thermodynamic system of specified character for which the term "thermodynamic entropy" should be reserved. The entropy given by the formula

$$S = -k \sum_n p_n \log_e p_n$$

in statistical mechanics, corresponds to the thermodynamic entropy, not in mathematical form but in the results it gives. The entropy of statistical mechanics describes a property of a <u>thermodynamic</u> <u>model</u> of specified characteristics which is selected on the basis of its appropriateness in portraying a <u>thermodynamic</u> <u>system</u>. Therefore, an amount of information may be identified with the negative of a change in thermodynamic entropy when the ensemble of interest in the information theory corresponds in behavior and constraint to the ensemble of the thermodynamic model employed to describe a system in thermodynamic equilibrium. Such a conclusion agrees substantially with that offered by MacKay (31):

> ..... when you define amount of selective information in terms of probabilities, you arrive at something which has the same form as the definition of entropy in statistical mechanics......we are prepared to operate in such a way that at our receiving end we can regard each of those signals as equally likely. Consequently, we are referring our question as to the amount of selective information to an ensemble appropriate to the assumption that all of those states are equally probable -- in which, if you like, all possible states are equally represented.
>
> When we calculate the amount of physical entropy, on the other hand, we are referring to the ensemble appropriate to a physical system in equilibrium at temperature T, for which not all possible states are equally probable.

[illegible]

[illegible]

$$[illegible]$$

[illegible]

[illegible]

[illegible]

..... And I think that all the debates and paradoxes which keep cropping up as to the relation between Shannon's amount of selective information and the concept of physical entropy disappear if one asks precisely what assembly is being used for the computation of the amount of selective information..... You get the physical measure if you use an assembly defined for thermodynamic equilibrium; and you get quite a different measure, of course, if you use the artificial assembly (the filing cabinet of the receiver) that regards all states equally likely. In that case, it is the metrical information content* and not the selective information content that correlates with physical entropy increase.
[Mackay 31]

* See Chapter V.

CHAPTER IV

SCIENTIFIC MEASUREMENT

1. Introduction

In the previous chapters, interest has centered on a mathematical definition of "amount of information" which has application in the study of communication systems. The implication has been that by arriving at a precise evaluation of the product of a communication system, i.e., an amount of information, and how this product is modified by noise and by the physical and probability constraints imposed, the investigator is enabled by suitable choice of system characteristics to achieve maximum efficiency. It has been indicated that equation (2.0) or its modification (2.15) are applicable in measuring the amount of selective information. The former is pertinent to the amount of semantic information and the latter to the amount of language information. In each we are concerned with statistical rarity -- of an event in (2.0); of a message in (2.15). The distinction between these measures lay in the specification of which representational ensemble, previously existing in the past experience of the receiver, was being employed by the sender. It was seen that the nature of the ensembles constituting a representation was also important in the comparison of entropy and information.

In the present and succeeding chapters, we are concerned not with the replication of pre-fabricated representations, but rather with the formulating of representations of some physical aspect of sensory

experience. The latter problem is treated in Scientific Information Theory. Here, as in the Communication Theory, the method is to make such definitions within the system so that the influence of the various components may be varied to optimize the functioning of the system for the intended purpose. In the following chapter, aspects of a Scientific Measurement system which correspond to various features of a Communication system will be proposed. However, in addition to a parallelism which appears reasonable, there is the relationship between information and measurement which has appeared in the Maxwell Demon discussions by Szilard and later by Brillouin. This relationship makes impossible the violation of the second principle of thermodynamics by the demon acting in a specified manner within a closed system. In order to effect a condition of lower entropy in the system, the demon requires information. However, since the demon obtains the information by a form of physical measurement, he produces an increase in the physical entropy (considering the entire system of demon, gas molecules, container, and measuring apparatus) of the system greater than the decrease he is able to accomplish with the information obtained. Brillouin concludes that the scientific experimenter is subject to the same type of restriction as besets the demon, and that there are limitations to the possibilities of measurements which have nothing to do with the uncertainty relations of quantum mechanics. [Brillouin 11]

Prior to a more detailed consideration of Scientific Information Theory in measurement, several viewpoints on measurement in general and on quantum measurement will be set forth. Such an endeavor will be brief and of limited selection as befitting the scope of this paper.

However, it will provide some insight as to how certain factors involved in Scientific Measurement might lend themselves to the methods of Information Theory.

Similarities between a communication system and a scientific information system may become more readily apparent in the following excerpts if it is assumed that the communication system of reference has the following features:

(a) A discrepancy between a signal as transmitted and a signal as received may be attributed to noise.

(b) Noise may enter the system at any point.

(c) Noise may be of the distortion variety in which there is a functional relationship between transmitted and received signals, or of the random variety in which there is no functional relationship, or of a combination of both varieties.

(d) Noise reduces the "amount of information received" in a message.

(e) Known constraints reduce the a priori uncertainty and hence reduce the amount of information received in a message.

(f) The transfer of information requires a transformation of energy.

(g) A message is a particular selection from an ensemble of possible messages.

(h) Manipulations upon information from a source tend to reduce the amount of information in a message. Translation or modulation from one code system to another or from one scale to another could be classed as manipulations in this sense.

(i) Ultimate delivery of a message to the human receiver necessitates that the message be put in a form which lies within his range of sensory response.

2. Scientific Measurement - General

Margenau (35) takes the position that the observer or experimenter is an entity but one that is in continuous interaction with the surroundings. He comments that failure to take into account the functioning of the observer within the system is outmoded and in disharmony with the successful phases of contemporary physics. There is common ground between the assumptions and rules of the scientist and the idea of constraints in communication theory. With regard to the former, Margenau states:

> Every scientist must invoke assumptions or rules of procedure which are not dictated by sensory evidence as such, rules whose application endows a collection of facts with internal organization and coherence, makes them simple, makes a theory elegant and acceptable. Ask an investigator why he prefers a simple explanation, why he hangs his knowledge of the universe upon a continuous and undifferentiated reference frame of space and time when his immediate experience is strongly accented by peaks of attention amid valleys of boredom.
>
> Now it happens that science in its more advanced stages is interested primarily in experiences of a highly specific type, called measurements. All measurements involve numbers. But this generalization should not be understood as barring from scientific interest many observations which do not yield numbers, examples of which are easy to cite. Suppose, for instance, that according to some theory a certain substance should emit a spectral line in a given spectral region and that according to another the line is forbidden. Whether or not it occurs is a matter of much importance, and it is settled wholly without an appeal to number. Again, it may be of great value to know whether two straight lines drawn on paper do or do not intersect. Observations of this sort again are not significantly represented as numbers; in our sense they are not measurements, but they are nevertheless important.

Turning now to measurements proper, we note a variety of ways in which they lead to numbers. Eddington believed that all measurements result from readings of the position of pointers on a scale, but in this he strained the facts for the sake of uniformity. To wit, there is at least one important kind of measurement that cannot be reduced to pointer readings, namely, counting. Much useful information was obtained by the early workers in the field of radioactivity through the tedious process of counting scintillations on a screen or by listening to the clicks of a relay activated by a Geiger counter. Observations on the growth of an embryo and on cell division yield numbers, though not via pointer readings. All these activities should be classified as measurements in the wider sense.

.....measurement involves (1) an object (in our terminology a physical system) upon which an operation is to be performed; (2) an observable whose value is to be determined; (3) some apparatus by means of which the operation can be carried out.

..... Spontaneous experience is richer than logic, to be sure, but it is also richer than language, which is a primitive form of logic. The rational can be adequately symbolized, either by ordinary language or in some other way, but the immediately sensed loses its fullness upon expression. Again the metaphor of a penumbra comes to mind. The process of translating experience into language may be likened to the projection of the shadows of objects upon a screen. A point source of light casts sharp geometrical shadows, a broad source surrounds each shadow with a region of haziness. It is as though the source of illumination increased in size as we proceed from reflective to spontaneous or sensory experience. We may now properly judge the transition from meaning to language to logic. Something vital is sacrificed in every one of the steps involved, and the loss is greatest in the field near perception. [Margenau 35]

Paraphrasing N. R. Campbell, we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules. The fact that numerals can be assigned under different rules leads to different kinds of scales and different kinds of measurement. The problem then becomes that of making explicit (a) the various rules for the assignment of numerals (b) the mathematical properties (or group structure) of the resulting scales, and (c) the statistical operations applicable to measurements made with each type of scale..... the most liberal and useful definition of measurement is "the assignment of numerals to things so as to <u>represent</u> facts and conventions about them." [Stevens 48]

In any measurement it is necessary to have some system that we regard as the measuring apparatus and from whose state we can draw inferences about the systems we are observing. In order that this be possible, it is necessary that the measuring apparatus interact with what is observed in a known and calculable fashion..... Hence, if we wish to make observations that are accurate enough to reach the quantum level, an element of incomplete determinism enters into the interaction between the apparatus and what is observed. This behavior is totally different from that predicted by classical theory, which says that the disturbance resulting from the measuring apparatus can be made arbitrarily small, and can be corrected for by means of the deterministic classical laws involved, even if it is not made negligibly small. [Bohm 6]

The whole subject-matter of exact science consists of pointer readings and similar indications. We cannot enter here into the definition of what are to be classed similar indications. The observation of approximate coincidence of the pointer with a scale-division can generally be extended to include observation of any kind of coincidence -- or, as it is usually expressed in the language of the general relativity theory, as an intersection of world-lines. The essential point is that, although we seem to have very definite conceptions of objects in the external world, those conceptions do not enter into exact science and are not in any way confirmed by it. Before exact science can begin to handle the problem, they must be replaced by quantities representing the results of physical measurement..... There is always the triple correspondence -- (a) a mental image, which is in our minds and not in the external world; (b) some kind of counterpart in the external world, which is of inscrutable nature; (c) a set of pointer readings, which exact science can study and connect with other pointer readings. [Eddington 17]

We have now to consider whether the doctrine that science must be based on observation needs any modification, in view of the fact that discoveries in physics of the most unexpected kind and of the greatest importance are frequently made by methematicians who have never performed, or even seen an experiment in their lives..... Before Maxwell, the story is one of performing experiments and devising formulae to represent the results. But the post-Maxwellian period is wholly different in character..... The change in the method of discovery after Maxwell may be illustrated by a simple analogy. Suppose that a map of Scotland is pasted on stiff cardboard and then cut up into small irregular pieces, so that it can be used as a jigsaw puzzle. Anyone who tries to solve the puzzle does not at first know what is <u>represented</u> and his only possibility of procedure is to find pieces which fit into each other and so constitute larger parts of the whole.

After a time, however, he will have progressed sufficiently to be able to guess that what is represented is Scotland, and from that time onward he completes the work not by finding pieces which fit into each other, but by using a priori knowledge of Scotland to put every fragment into its proper place. These two methods may be likened to the two types of research in physical science: the earlier, proceeding step by step by experiment in special topics; and the later, knowing a priori what ought to be, because a guiding principle is now available for the whole, permitting extension of knowledge by purely rational methods. [Whittaker 54]

The instruments of thermodynamics include thermometers and instruments for determining the various mechanical parameters, such as pressures or stresses or electrical or magnetic fields. They must not be large compared with the geometry of the boundaries of the systems we have to deal with, and they must be small enough so that with the help of them the given system can be analyzed into elements each of which is sensibly homogeneous..... As the size of the instruments is diminished, the data first pass through wide fluctuations imposed by the gross geometry of our system; that is, at first a single instrument may be trying to straddle a piece of iron and a piece of copper. As the instruments get smaller, their indications smooth out and approach a smooth level plateau. As they get still smaller, fluctuations again begin to manifest themselves. The universe of thermodynamic operations is restricted to the region of the plateau. It is also a matter of experiment that there is such a plateau. [Bridgman 7]

Dimensions and Measurement

(1) A physical quantity may be taken as anything that can be measured by one or more strictly definable processes. (2) A measurement of a physical quantity generally consists (and could, if desired, always consist) in principle of a numerical comparison of the quantity with an arbitrarily chosen unit; the result of the measurement, represented in physical equations by a symbol, will be called the magnitude of the corresponding physical quantity. (3) Every magnitude appearing in a general equation represents the result of a measurement of a physical quantity by a unique, strictly specified process. (4) Magnitudes are of two kinds -- fundamental and derived. A fundamental magnitude is one whose value is unaltered by any change in the process of measurement, or in the chosen unit, of any physical quantity other than the one to which it refers. A derived magnitude is one whose value is in general altered by such a change. (5) The number of fundamental magnitudes is arbitrary. (6) A derived magnitude may be uniquely expressed in terms of

> those fundamental magnitudes, a change in the units or processes of measurement of which produces an alteration in its value. (7) Physical equations are of two kinds — definitions of derived magnitudes, and experimentally established relations. (8) When every magnitude occurring in a physical equation is reduced to fundamental magnitudes, every term in the equation consists of the same magnitudes raised to the same powers; i.e., the equation is homogeneous in each magnitude. (9) The power to which each fundamental occurs in the reduced expression of a term in a physical equation is called the dimension of that fundamental magnitude in the corresponding term. (10) Dimensions are characteristics of magnitudes, which are the results of measurements of physical quantities by strictly specified processes; they are not characteristic of physical quantities themselves. [Dingle 16]

Dingle, in commenting on ambiguity in modern physics concerning the choice of fundamental magnitude, gives two examples: the measurement of temperature and the measurement of time in which incompatibilities result because of absence of agreement on how these magnitudes are to be measured.

> ..... As things are at present, one often does not know what is meant when certain magnitudes are mentioned. This would be bad enough if only formal expressions were at stake, but actually matters are much worse; it is the laws which our equations express that have become ambiguous, and the ambiguity is not realized..... When length is measured in terms of a standard rod, and time in terms of the rotation of the Earth, Newton's First Law of Motion becomes a hypothesis to be tested. One test (indirect, but valid) is the comparison of the observed with the calculated tracks of ancient eclipses. A discrepancy is found, which means that Newton's law is inaccurate.
>
> Astronomers, however, do not draw this conclusion; they say that the Earth is slowing down, while Newton's law remains true. But this means that the rotating Earth is abandoned as the standard of time measurement, and the scale implied by Newton's First Law is substituted for it. Equal times are then, by definition, those in which an undisturbed body moves over equal lengths, and the approximate uniformity of rotation of the Earth becomes a fact of observation. A still further change was made when Einstein substituted a beam of light for an undisturbed body; the "postulate of the constancy of the velocity of light" is,

in effect, a definition of the scale for measuring time. Tacitly adopting this scale, Eddington can say, "my personal conclusion is that there is no more danger that the velocity of light.....will change with time than that the circumference-diameter ratio pi will change with time." Such a conclusion is possible only if light defines the time-scale, but by stating it as a "personal" conclusion, Eddington gives the impression that it is conceivably false. The active existence of two incompatible time scales in physics is thus clearly seen..... Let us now see how the dimensions of time are affected by this duality. If time is measured in terms of the rotating Earth, it is a fundamental magnitude, and if dimensions are simply (T). If, on the other hand, it is measured in terms of the space covered by a moving body or by light, it is a derived magnitude, for a change in the measurement of length would make a change in the value of a time magnitude. The equation (choosing light instead of an undisturbed body, for example, and choosing 1 cm. as the distance defining the unit of time) is

$$t = \frac{1}{3 \times 10^{10}} \ell$$

whence t must have the dimensions (L). Hence again we have incompatible definitions yielding different dimensions; and until we decide how time is to be measured, we cannot assign dimensions to any magnitude derived from time.

3. Quantum Measurement

Any consideration of measurement must also include views on quantum measurement.

Atomic events manifest themselves by their ingression into macroscopic experience. The methods we have described of investigating the properties of atomic systems exploit this continuity between atomic and macroscopic events. Through the observable effect of photons on a photographic plate and through the observable increase in the energies of photoelectrons, we are able to extend the concepts of position and energy to photons. In a sense, we perform measurements on atomic systems when we investigate them in this way. Now, every measurement disturbs the system being measured. Classical physics rests on the supposition that all measurements on a system can be performed so gently that the disturbances they cause are negligible. The quantization of radiant energy, indicates, however, that there is a lower limit to the disturbances caused by the most gentle measurements -- those employing the interaction of light with the system. Thus, even if we regard atomic systems as geometrical configurations, our measurements will disturb these systems in an essentially unpredictable way. [Menzel 36]

Menzel then points out that Bohr and Heisenberg demonstrated convincingly that the indeterminacy relations could be thought of as arising from the unpredictable nature of the disturbances incident on measurement. Also that Von Neuman, having interpreted the mathematical formalism of quantum mechanics in the light of the above ideas, showed that the changes in a system resulting from a measurement on it are irreversible in the sense of the second principle of thermodynamics. Menzel objects, in concluding, to attributing the uncertainty solely to the measurement. He prefers to include the very nature of the atomic order in the explanation. He regards the atomic order as positive and objective but, nevertheless,

> .....we must recognize that the atomic order, because it is formulated in terms of physical quantities, is an order that depends on measurement. Here we use "measurement" to mean the "methods by which we extend the meaning of physical quantities to apply to non-macroscopic systems. [Menzel 36]

Bohm (6), commenting on an attempt to avoid the difficulty of an unpredictable and uncontrollable transfer of a quantum in the interaction between observing apparatus and what is observed, by considering the observing apparatus and what is being observed as part of a common system, states:

> The chief difficulty with the procedure outlined above is that it yields us no information. In order to obtain information from the system, we must interact with it somewhere, for example, by looking at the photographic plate, and in so doing, we will have to use light..... Thus, when we use the plate in such a way as to provide information about the position of the electron, we inevitably make the momentum of the combined system (camera, plus plate, plus electron) indefinite.
>
> In all cases, one obtains information by studying the interaction of the system of interest, which we denote hereafter by S, with the observing apparatus, which we denote by A. Any object whose properties are understood, even if only in part, can in principle by utilized in the construction of the observing apparatus. Although every observation must be carried out by means of an interaction, the mere fact of interaction is not, by itself, sufficient to make possible a significant observation. The further requirement is that, after interaction has taken place, the state of the apparatus A must be correlated to the state

[illegible]

[illegible]

[illegible]

of the system S in a reproducible and reliable way. This correlation is in general statistical, but in limiting cases it may approach any conceivable degree of exactness..... Thus, in a typical observing apparatus we obtain a correlation such that each clearly distinguishable state of the apparatus corresponds to a range of possible states of the system under observation. This range may be called the uncertainity, or the error, in the measurement. The possibility of error usually arises from defects or inadequacies in design of the apparatus that are, in principle, avoidable. In extremely accurate measurements, however, it may arise from the quantum nature of matter, in which case a more accurate measurement cannot be made without changing what is observed in a fundamental way.

Bohm further points out that all real observations are, in their last stages, classically describable.

We may give as an example the usual practice in science, whereby one obtains data from meter readings, spots on a photographic plate, clicks of a Geiger counter, etc. All these objects and phenomena have the common property of being classically describable. A little reflection will convince the reader that all observations ever made in science have employed at least one such classically describable state..... If the investigator wishes to study the quantum properties of matter, he requires apparatus that amplifies the effects of individual quanta to a classically describable level..... If a sharp distinction could not be made between the observer and the systems observed, scientific research as we know it would not be carried out, because the observer would not know which aspects of an observation originated in himself, and which originate in the outside systems of interest. We do not wish to imply, however, that scientific research is necessarily impossible whenever an observer interacts significantly with the things that he observes; for as long as the observer can correct for the effects of his interactions, on the basis of known causal laws, he can still distinguish between effects originating in him and those originating outside.

.....a measurement process is irreversible in the sense that, after it has occurred, re-establishment of definite phase relations between eigenfunctions of the measured variable is overwhelmingly unlikely. This irreversibility greatly resembles that which appears in thermodynamic processes, where a decrease of entropy is also an overwhelmingly unlikely possibility. Because the irreversible behavior of the measuring apparatus is essential for the destruction of definite phase relations and because, in turn, the destruction of definite

phase relations is essential for the consistency of the quantum theory as a whole, it follows that thermodynamic irreversibility enters into the quantum theory in an integral way. This is in remarkable contrast to classical theory, where the concept of thermodynamic irreversibility plays no fundamental role in the basic sciences of mechanics and electrodynamics. Thus, whereas in classical theory fundamental variables (such as position or momentum of an elementary particle) are regarded as having definite values independently of whether the measuring apparatus is reversible or not, in quantum theory we find that such a quantity can take on a well defined value only when the system is coupled indivisibly to a classically describable system undergoing irreversible processes. The very definition of the state of any one system at the microscopic level therefore requires that matter in the large shall undergo irreversible processes. [Bohm 6]

Speaking within physics rather than philosophizing about it, we use the term "measurement" very broadly. We say that we measure the temperature of a gas, but we also say that we measure the (average) velocity of its molecules. These are two different things. The difference I have in mind is not that in the first case we simply read an instrument, while in the second we derive the numerical value from several such readings through a fair amount of computation. The important difference is, rather, that in the case of temperature we measure an empirical construct, while the second number receives its full meaning or interpretation only as an additional step, the coordination of, say, the classical kinetic model to the empirical constructs and laws of thermodynamics. Measurement (in terms of immediately observable empirical constructs) is based on the observation of scales, and I have never heard it suggested that we make a needle move by watching it, which is but another way of saying that on the common sense level of laboratory objects and their immediately observable properties and relations, the language of commonsense realism is the only reasonable one..... In measuring an empirical construct exemplified by an object or situation A at a given moment -- or, as I shall say briefly, in measuring A -- one does not observe A alone but, rather certain aspects of a situation (A, B) compounded of A and the yardstick or measuring instrument B. There is thus the possibility of an interaction by which the two components of the new situation, A and B, may produce changes in each other. That gives rise to two questions: (i) how can we recognize such changes? (ii) under what conditions is a feature of (A, B) acceptable as a measurement of A, that is, as an index or characterizer of A alone? [Bergmann 4]

The answer to the first question is self-evident. We shall say that A has been changed by being put in the measuring situation if it subsequently behaves in some respect differently from A' -- which is otherwise exactly like A, but has not been measured -- provided that the difference cannot be attributed to other factors. If the differences occur only while (A,B) is maintained, the change may be called temporary... .. A property of (A,B) is a measure of A if and only if it enters, together with other such properties of A (and of other objects), into empirical laws that predict or postdict the behavior, before or after the occurrence of (A,B) of A (or of A interaction with other objects)..... One may measure the length of an iron rod with an ordinary yardstick to the nearest full inch, or one may measure the same stick with a more elaborate instrument to the nearest 0.01 in. In either case, as in all measurement, one manipulates physical objects and, eventually, reads a scale. The perceptual exertion required may actually be greater in the first case than in the second. Yet we call the second measurement more precise than the former -- or this, at least, is how I shall define 'precision.' Precision, then, means the number of digits of a given unit. The larger this number, the greater the precision. How precise we can be is a matter of empirical laws and, in particular, of those empirical laws that are sometimes referred to as the theory of the instrument. On the other hand, a measurement whose precision is much less than the best we can do may be completely reliable, a measurement being called reliable when in a large number of repetitions the result is always the same.

.....If the necessary care is taken, the first of the two measurements of the iron rod is, in fact, completely reliable. The second measurement which is more precise, is less likely to be completely reliable. The values obtained will scatter or, as one also says, their standard error will not be equal to zero. Having thus defined precision and reliability, I turn to a definition of accuracy. The following is, I believe, an exact statement of that rather fundamental feature of our world to which we refer when we say that there is, in fact, a limit to the accuracy of our measurements. A measurement as precise as we can make it is never completely reliable. Its standard error, through absolutely decreasing with increasing precision, shows no tendency to decrease in proportion to the last digit. Conversely, if our most precise measurements were completely reliable, we would not consider them as of limited accuracy..... As is well known, we do not in careful experimental work expect our measurements to be reliable. We repeat them, define their average as the "true value" and operate in the formulation and testing of laws with the value thus obtained. Anybody who wishes to describe this state of affairs by saying that all laws of nature are "statistical" is free to do so.

But having made this choice of meaning, he is no longer free to use the same term in a different and more specific sense in which not all but only some empirical laws and theories are statistical. Or, at least, he may not do so without being explicit about it. Furthermore, anybody who is thus explicit will not be tempted to believe that the inaccuracy of measurement, by making all laws "statistical," implies or even suggests the statistical "nature of the quantum theory." [Bergmann, G. 4]

The most nearly complete information obtainable about a quantum-mechanical system is summarized in its state. But this state is not itself the object of physical measurement. As a matter of fact, most measurements on a system change its state in an unpredictable fashion..... Bohr has taken the attitude that the fault of classical physics lies in that it attempts to discover physical reality in one object taken in isolation and that, as a result, causality and reality tend to evaporate before our eyes. He suggests that we should consistently look at the physical object <u>and</u> our measuring devices as the unit to which causality and reality must be applied..... Einstein, one of the early workers in quantum physics, has consistently held that quantum mechanics is a temporary state of the theory, which must be overcome ultimately by a theory that resembles classical field theory much more closely than it does quantum mechanics. Though agreeing that in any observation we make, our measuring equipment interferes with the objects we wish to observe he feels that in our theoretical description we ought to be able to conceive of the object apart from its interaction with the measuring instrument. [Bergmann, P. G. 5]

Bergmann then concludes with a statement of his own position:

Our physical measuring instruments consist themselves of the same basic ingredients as the rest of the universe, and I do not believe that the interaction between a measuring device and the object to be measured is different in principle from the interactions of any other physical objects. Whether we care to read a dial or not, in other words whether we complete the observation or let the measuring instrument remain part of the unobserved universe, cannot affect the behavior of the instrument. On the other hand, quantum mechanics shows that in general we lack sufficient information concerning the initial relationship between object and measuring device to predict with certainty the result of the interaction. It is possible to construct exceptions to this general rule, however, just as in particular situations it is possible to predict the outcome of measurements..... Thus it would appear that at least some aspects of the wave function of de Broglie and Schrodinger contain the "reality"

of a physical situation, but there remains the question whether we can analyze more precisely the effect of a measurement on this wave function than is usually done. My point of view would seem to lie somewhere between those professed by Bohr and by Einstein, but probably closer to Einstein's.
[Bergmann, P. G. 5]

Quantum mechanics gives a very clear and unique answer to the question as to which possible results we may expect when we measure a certain observable, represented by an operator with certain eigen-values. We get an equally clear answer if we ask how great the probability of one of the possible results will be, provided a definite "state" or wave function is given. But there remain some questions about the process of observation itself -- questions for which we do not get unambiguous answers because orthodox quantum mechanics treats the concept of "measurement" as a fundamental one which ought not to be analyzed. It is not so clear, however, whether this attitude can be maintained without exceptions or restrictions..... But while thermodynamics is essential for the concept of observation and measurement, this concept itself seems to me to be indispensable in thermodynamics and in the notion of entropy. The relations of thermodynamics and quantum mechanics - especially thermodynamical statistics and quantum mechanics - has been the object of much discussion. Let us mention here only the first and last stages of the subject. (1) Pauli emphasized that even in quantum theory there remains the necessity of an "hypothesis of elementary disorder," which has to be acknowledged as an additional axiom besides the "pure" quantum mechanics as formulated by the Schrodinger equation..... (2) During the last years, Born and Green, in a series of papers, developed a fascinating account of thermodynamical statistics based upon quantum mechanics. Those results of their endeavour which are related intimately to our question here may be formulated in two theses: (A) Quantum mechanics in its full content implies irreversibility as a necessary consequence, but (B) "pure" or "restricted" quantum mechanics, which applies only to the Schrodinger equation without the concepts of preparations of states, observations, measurement or "decision," would not do so.
[Jordan 27]

(Speaking on Bohr's principle of complementarity, Oppenheimer has stated the following:)

The basic finding was that in the atomic world it is not possible to describe the atomic system under investigation, in abstraction from the apparatus used for the investigation, by a single, unique, objective model. Rather, a variety of models, each corresponding to a possible experimental arrangement and all required for a complete description of possible

physical experience, stand in a complementary relation to one another, in that the actual realization of any one model excludes the realization of others, yet each is a necessary part of the complete description of experience in the atomic world.
[Oppenheimer 36]

4. Summary

Similarities between the communication and measurement systems suggested by the excerpts presented are:

(a) General

1. The assumptions and rules of the scientist may be likened to constraints.

2. Ultimately the observer obtains the information on a sensory level.

3. Measuring apparatus between the object and the observer correspond to modulators which operate on inputs from the object or phenomena, or on outputs from other modulators.

4. Since the object can contribute to a representation in the observer, it may be considered as a source of information.

5. Modification of the output of the source tends to increase in passing from the source to the observer through the system of apparatus.

6. Errors in measurement might be compared to noise effects.

7. The a priori and a posteriori states of the observer and the system of apparatus and the system under investigation must be considered in evaluating the amount of information.

8. The change in the system measured, produced by interaction with the system of apparatus, corresponds to improper receiver modulation.

CHAPTER V

SCIENTIFIC INFORMATION THEORY

1. Introduction

Previously it has been stated that communication theory is concerned with the problem of reproducing a representation which already exists somewhere else and that Scientific Information Theory is concerned with the problem of formulating a representation of some physical aspect of sensory experience. In the preceding chapter, background material was presented to suggest that the problem of formulating a representation had many similarities to the problem of relicating a representation. It was noted that the obtaining of information by means of physical measurement is accomplished at the expense of an overall increase in the entropy of the system made up of phenomena of investigation, system of apparatus, observer, and the environment. In considering the communication problem, a definition of the amount of information was given which was suitable for a mathematical study of information from the standpoint of selective information. Here, in the Scientific Information System, definitions of the amount of information applicable to the nature of this sytem will be given which are also suitable for mathematical study. In the scientific information system, we shall be interested in the amount of <u>structural information</u> and the amount of <u>metrical information</u>. The discussion which follows is based upon a theory of scientific information proposed by MacKay (33,34). It should be noted, however, that the result of a measurement may be

[illegible]

considered from the standpoint of the amount of selective information; this measure of the amount of information should be distinguished from those now discussed.

2. Structural and Metrical Information

Measures of information are the <u>structural information content</u> and the <u>metrical information content</u> [which is related to Fisher's "amount of information" (20)]. The distinction between these measures can be illustrated by considering a typical expression of the result of a scientific measurement. "Value X corresponds to interval Y." Structural information is concerned with Y; metrical information is concerned with X. In the design of experimental apparatus and procedure the observer is enabled to formulate certain distinguishable and independent "blank statements" or propositional functions a priori. The actual experiment then consists in obtaining evidence with which to fill in the "blank statements." The problem here, then, is the operational definition of Y and the collection of evidence for X. The structural information content may be defined as the number of independent propositional functions which we are enabled by a particular experimental method to formulate. This could be described as the number of logically distinguishable degrees of freedom of the representation. Each of the blank statements mentioned above signifies one independent respect in which the representation could be different. Thus, as in the case of communication aspect, information theory proposes a more explicit method for considering problems of scientific information by statistical and analytical techniques. Units are defined for the structural information content and the metrical

[illegible]

information content -- the "<u>logon</u>" and "<u>metron</u>" respectively. (See Appendix). The information content of a given representation is specified by setting down the metron content of each logon. Analysis of the information content is facilitated by employment of an "information vector space" or of matrix algebra, neither of which will be considered here. An example will aid in understanding the general notions of structural and metrical information content. Suppose that it is desired to represent the voltage of a signal coming through a channel of a certain band width, as a function of time. At certain intervals, we want to take "new" readings to provide "new" ordinates for a graph. If the readings are taken too close together, they are practically the same reading since the inertia of the system prevents very rapid changes. Gabor has shown that in the ideal case there is a minimal separation in time between readings, below which (according to a certain criterion of independence) they cease to be "practically independent." This minimal separation, $\triangle t$, is related to the band width, $\triangle f$ by a relation of the form

$$\triangle f \cdot \triangle t \geq K \tag{5.0}$$

where K is a constant depending on convention, but of the order of 1/2. Thus in the time t, apparatus with a band width f enables one to formulate about 2 X f X t independent propositions about the signal amplitude. Here then is a measure of the number of labels or "blank" statements which the experimental method provides before performing the experiment. It is the structural information content of the ultimate description of the signal. The <u>metrical information content</u>

in this instance can be measured by

$$\left(\frac{V}{N}\right)^2$$

where V is the voltage amplitude and N is the noise amplitude. The variance is the square of the noise amplitude; thus the connection with Fisher's "amount of information," which in the simplest case is measured by the reciprocal of the variance of a statistical sample. Without defining logon or metron we shall briefly discuss their implications.

(a) Structural Information

When a chain of apparatus is involved (including the observer), then the differentiating capacity of the least-discriminating link determines the logon content (number of independent categories) in the result. In many cases, structure is defined in terms of a reference-coordinate. For example the density pattern on a photographic plate can be described by a function of one or more space-coordinates, and the structure of a telephony signal can be specified by a time function. The logon-capacity of an experimental method can in such cases be defined as the number of logons which it specifies per unit of coordinate-interval, or coordinate-space if several coordinates are involved. Thus the logon-capacity of a microscope in a particular region in the focal plane can be defined in logons/cm^2, and measures the resolving-power in that region. The logon-capacity of a galvanometer or a communication-channel is measured in logons per second, and represents the number of (practically) independent readings per second which can be made with the apparatus. The logon-capacity

[illegible]

of an instrument is related to its frequency bandwidth, where the latter is defined for the general case as the effective range of input-frequencies to which it is sensitive, by the relation

$$\Delta f \cdot \Delta q \geq K_s \tag{5.1}$$

where Δf represents the effective range of frequencies (conjugate to a coordinate q) to which the apparatus is sensitive, Δq twice the uncertainty in q, and K_s a number having value about 1/2.

To attempt to talk of "an interval smaller than Δq" would be to try to construct a logical pattern identical with that of "a frequency bandwidth greater than Δf" which cannot by definition appear in any result and is therefore observationally meaningless. It is interesting to note that the uncertainty relation of quantum mechanics

$$\Delta E \cdot \Delta t \geq \frac{h}{4\pi}$$

which is similar to equation (5.1) may be considered as a measure of the absolute logon-content in Quantum Theory.

(b) Metrical Information

The quantal character of metrical information arises from the way in which a scientific measurement is described. [Mackay] A description of a result is basically a set of instructions enabling the reader to reproduce for himself the conceptional pattern representing the experience of the observer. The most elementary observational proposition asserts the existence of a coincidence-relation between two entities. On the other hand, a magnitude is defined by saying that it occupies a certain interval on a scale. Logically this occupance-relation

[illegible]

between scale-interval and magnitude is a consequence of the existence of coincidence-relations between the ends of the "unknown" and two definable graduation-entities on the scale. For every observation there is a minimum separation between neighboring graduation-entities below which either we cannot define or cannot substantiate with probability greater than one-half, a proposition of the form: "A falls into B_{n-1} -- B_n and not into B_n -- B_{n+1}. Thus what we carry away from a measurement is basically an integer, the number of conceptually separate occupance-relations which have been specified. This integer is concerned with the metron-content of the result. The metron content of a result must be incapable of augmentation by purely logical manipulation, and all complete representations of a given result should have the same metron-content.

The quantization of scientific information according to the definitions of structural information content and metrical information content just discussed is amenable to mathematical analysis and hence is an aid in the study of experiments or of a scientific information system. A statement of the result of a scientific measurement may be regarded as a complex of the quanta of structural and metrical information. Thus the abstraction from scientific statements related to measurement of a logical form which is quite general leads to a clarification of experimentation, the role of the experimenter, and of fundamental relations in the different fields of physical science. Mackay (33) has expressed the need for such a general tool of expression as follows:

> Experimentation abounds with indications that the everyday concepts of science are not the most fundamental. Each time that a compromise has to be struck, say, between the sensitivity and the response-time of a galvanometer, or the noise-level and band-width of an amplifier, or the resolving power and aperature of a microscope, one has an intuitive feeling that in each case some quantity is remaining constant behind all experimental manipulations -- something more fundamental than either of the quantities in question. We say that Nature cannot be cheated; and examples of this principle recur throughout the realm of measurement, and not only in microphysics.
>
> Is there not then a way of expressing scientific facts so that in any context a single universal principle can apply? Presumably in sufficiently fundamental terms such a principle should become obvious.

It is interesting to note that the compromise which must be accepted in a scientific measurement system is similar to the compromise which must be accepted in a communication system between band width and noise level. The latter concept has received considerable attention in the study of communication systems by the methods of information theory. Another compromise is apparent in the demon problem where information is gained at the expense of increasing the total entropy. With this in mind we shall consider the role of entropy in Mackay's Theory of Scientific Information.

3. Entropy and Scientific Information

At this point it will be well to restate the relation between selective information and entropy before considering the applicability of the latter to "scientific information." Selective information content may be <u>identified</u> with the entropy of statistical mechanics in the particular case where the ensemble from which the selection is made is a physical one defined for a state of thermodynamic equilibrium. In this instance "information" will be measured in units of ergs per

degree centigrade. An alternate procedure would be to measure temperature thermodynamically in work units so that Boltzmann's constant would take the dimensions of a pure number related to a thermodynamic system in equilibrium. If this is done, then both "information" and the entropy of statistical mechanics will appear as dimensionless numbers but still related to the thermodynamical equilibrium ensemble in question. Whenever we are discussing thermodynamic entropy by the methods of statistical mechanics, we must keep in mind that in order to study the properties of a _thermodynamic system_, whose condition is described by the values of a limited number of thermodynamic variables, we must consider the average properties of an appropriately chosen _representative ensemble_ of systems, of similar constitution to the one of actual interest. In a general way it may be said that the appropriate choice of representative ensemble depends on taking a distribution of the members of the ensemble over their possible individual states, which agrees, on the one hand, with our knowledge of the thermodynamic variables that have been measured, and which conforms, on the other hand, with the hypothesis of equal _a priori_ probabilities and of random _a priori_ phases on which the deductions of statistical results have been based. The condition of thermodynamic equilibrium for the systems of usual thermodynamic interest can best be represented by a _canonical ensemble_, since this has been found to give the most appropriate description of equilibrium in the case of systems in thermal contact with their surroundings or in essential rather than perfect isolation therefrom. Hence for "information" (selective) to be identified with physical entropy its ensemble must be limited in a

like manner. Frequently the mathematical form $-\sum p_i \log p_i$ (the entropy of a probability distribution) which is a pure number and which appears in the H-theorems and in statistical mechanics, has been labeled "entropy." If entropy is considered to be defined by this expression, then it is readily identified with selective information. But even here this form is related to an ensemble with certain constraints and with definite properties. Thus one should determine the nature of the system and the ensemble in question before identifying selective information content with the statistical mechanical analogue of thermodynamic entropy.

With regard to "scientific information," Mackay states that the <u>metron-content</u> of a measurement and the entropy are equivalent quantities, both having quantal aspects, and a change in one being opposite in sign to the change in the other. Thus in a physics which started from the concept of Information as one of its basic quantities, the sum Entropy-plus-Information content would rank as a fundamental invariant.

A system whereby a representation is defined by a selection process is termed a <u>code</u> <u>system</u>. The corresponding representation of the selection process transmitted is known as a <u>code</u> <u>signal</u>. As a physical sequence the code signal itself will have metrical and structural features and will be definable by a vector in an information space. BUT ITS STRUCTURE NEED NOT HAVE ANYTHING IN COMMON WITH THAT OF THE <u>REPRESENTATION</u> WHICH IT IDENTIFIES (the tip of the information vector occupies one of a number of cells into which the information

space is divided or quantized; the ease with which one of these possible positions is selected by the receiver need not be dependent upon how the information vector was built up by logon and metron contents). However, the ordinary case of making physical representations in Scientific Information Theory, Mackay (33) could be thought of formally as a special case of coding, one-for-one. Thus the result of an experiment, as well as a communication signal could be analyzed in terms of its selective information content. This is a relative measure, depending on the number of distinct results which were regarded as equally probable by the observer. The result observed is thought of as specifying one of a number of possibilities already contemplated by the observer as forming an ensemble in defined proportions. The amount of selective information derived from the experiment can then be computed in the same way as for a message. Therefore, it is apparent that the information content of an experiment can be determined from the selective, or from the structural and metrical viewpoints, or both, depending upon whether our interest lies in the question, "How unusual or unexpected is it?" or "How big is it?" or "How much detail has it?"

Again it should be emphasized that regardless of which viewpoint is chosen to measure the information content of an experiment, that the selective information content may be identified with the thermodynamical physical entropy only in the particular case where the ensemble from which the selection is made is a physical one defined for a state of thermodynamic equilibrium. If all n distinguishable voltage-levels of a transmitted signal are regarded as equiprobable, the selective information content per logon is proportional to $\log n$.

[illegible]

On the other hand, the physical entropy increase is proportional to or must exceed n^2. Here the correlation is between metrical information content and physical entropy increase. Metron content can be thought of here as the number of unit increases of physical entropy -- i.e., of elementary events -- which have been subsumed under one head, thereby losing their distinguishability and potentiality of serving as "bits." Under optimum conditions, the energy change is a minimum; and this, in general, is proportional to the amount of metrical information.

(a) Example Problem

On the basis of material presented in this and previous chapters, and in order to show the significance of some of the ideas proposed, the following simple empirical problem is presented. Although not practical in itself, the situation is devised to illustrate how an experimenter might apply concepts of information theory to explain measurement phenomena.

The materials used, statement of the problem, and requirements are as follows:

(a) Materials used:

(1) Two large baths containing equal volumes of ice and water in equilibrium, both volumes having been drawn from the same initial container. The two baths are insulated such that both volumes will remain at identical temperature if allowed to continue isolated from the exterior.

(2) Two small baths containing minute though equal quantities

of ice and water in equilibrium drawn from the same initial container used in (1). Further restrictions are exactly as stated in (1).

(3) One high heat capacity thermometer at room temperature (25 Degrees C) with bulb of such dimensions as to be small enough to be adequately covered if inserted in small volume described in (2).

(b) Statement of the problem -- measure with above thermometer one each of volumes (1) and (2).

(c) Requirements -- report temperatures obtained and if not identical or reasonably close, determine which is correct and why.

We will assume that the experimenter's starting point is one of the small volumes, and that he does not know that the temperature should be very close to 0 degrees Centigrade depending upon the accuracy of the thermometer.

The experimenter, striving for accuracy, makes several measurements of the small volume and attains readings ranging from 6 to 10 degrees with the mean at 9 degrees Centigrade. Knowing beforehand that all volumes came from the same initial source, and that therefore the large and small volumes should have approximately the same temperature distribution, our experimenter turns to one of the large volumes. On the basis of his measurements just completed and the previous statement, he forms a mental picture or representation of what the results should be for his next run, a priori assigning probabilities for specific readings. However, upon making his measurements, he

finds that the readings from the large bath all range in the vicinity of 0 degrees Centigrade. Since these indications fall on the border or even outside his proposed pattern, his first conclusion might be that he has received a large amount of information.

Yet, because of the discrepancy, that conclusion does not quite satisfy him. According to information theory as applied to physical systems, there should be a large entropy change accompanying a receipt of much information from a measurement process. The experimenter, upon turning his attention to the large and small vats which were unmeasured, notices no visual change if he compares large volumes. However, the comparison of small volumes does indicate differences. The ice content of the one whose temperature was measured seems to be less than that of the alternate one. He believes it possible then that there might have been an interaction between his measuring instrument and the measured small volume, and that, as a consequence the results attained therefrom portray an erroneous picture. Since this picture was applied to give a priori probabilities to his representation of the large volume, it could be the reason why he seemed to get so much information from the large volume when there was no apparent change in the bath. The experimenter's original picture, in that case, should have been a close approximation to the final results, and, therefore, little information ought to have been received. He decides that the results for the large bath are more nearly correct.

The initial measurement of the small system was undertaken with no representation in mind. If the experimenter considers that system again, the above results gained from the large bath then should determine

his a priori pattern of the small system. Since the indications of the measuring instrument diverged to some degree from this pattern, and the "source of information" was disturbed, the experimenter really received much "information" from the small system. However, this he cannot consider as "good information," rather he now knows that he was deceived by the receipt of misinformation in a form of distortion noise introduced by interaction between the measuring device and the bath. If he now selects a thermometer of small enough dimensions such that the bulb contacts only a slight amount of the mixture, he can approach more closely the distribution attained by the measurement of the large volume completed above and deemed correct.

The foregoing approach to this problem has been based solely on the observer's forming a representation based on preconceived possibilities. His results then are all in the realm of selective information. Earlier in Chapter V, reference was made and explanation proposed with regard to another means of attacking the problem of measurement. This means is concerned with "logon content" or structural information content and "metron content" or metrical information content. What results are yielded by the application of these concepts to the above example?

The experimenter here is concerned with only a single logon, that of temperature. This is true because fluctuations can arise from only two sources: (a) those due to the random collisions of the molecules with the thermometer, and (b) those arising as a result of the gradual change over a long period of time of the system. The former are of such rapidity that they are unobservable in any given temperature

reading because of the design of the instrument. Upon observing equation (5.0), it can be seen that for the usual time required to take a temperature reading the latter fluctuations will not affect the results because of their low frequency.

In the absence of information as to the temperature, the experimenter might assume all temperatures measurable with his device to be equally probable. If in measuring the small volume the experimenter takes several readings, and between readings allows the thermometer to return to its initial state, his results will show a decided spread. The variance given by σ^2 will then be of considerable magnitude. Since metrical information content is related to the reciprocal of the variance through Fisher's measure, this quantity will be low.

On the other hand, similar measurement of the large volume will yield results all of which should be of nearly the same magnitude, thus having little spread, and the metrical information content will be high.

If he considers the signal to noise ratio, with the understanding that the variance is the square of the noise amplitude, the cause of the discrepancy arising between the measurements of the two systems becomes clearer. The small volume results having a larger variance must someway have been subject to quite a large amount of noise. The large volume results do not show having had the same effect introduced. Thus, since the metrical information content is higher for the large volume, the experimenter can have, in effect, more confidence in his results obtained there.

Thus by the application of two different concepts of "information," that of selective information and that of metrical information, the

observer has been able to determine unforeseen difficulties arising when he operated on the system in order to learn one of its gross characteristics.

4. Efficiency Determination by Information Theory

In considering entropy and scientific measurement or experimentation, Brillouin analyzed "observations" in terms of the Selective information gain and the physical entropy cost of the "observation." The relation involved is expressed in terms of the efficiency of the experiment

$$\mathcal{E} = \Delta I / \Delta S_0 \leq 1 \qquad (5.2)$$

This physical entropy cost corresponds with the metrical information content which has been identified with the minimum physical entropy change produced by the measurement itself under optimum conditions. Szilard's demonstration of the validity of the second principle despite the operation of Maxwell's demon* in a system has indicated a generalized statement of the second principle for any process in which physical measurement is involved of the form

$$\Delta(S_0 - I) \geq 0 \quad , \quad \text{where} \qquad (5.3)$$

S_0 = initial physical entropy I = selective information

Thus a measurement in which the selective information gain was low relative to the metrical information content would be one of low efficiency. The entropy increase introduced by the measurement is related to the monetary cost of conducting the experiment and to

* See Chapter IV, Introduction.

subsequent measurement of properties of the higher entropy system. Thus it is worthwhile to consider whether or not the selective information gain is compatible with these factors and to make optimum use of the system and techniques available to increase efficiency.

After a consideration of a number of examples, Brillouin (13) concludes that for measurements of high accuracy, the efficiency according to the above definition could be low; if extremely small distances had to be measured, the efficiency of the observation could drop to 10^{-15}. Thus he states:

> The physicist operating in a given laboratory disposes of a limited supply of negentropy, which results in a limit to the small distances he can actually measure..... The conclusion is that there is no precise limitation to the small distances that can be measured but that the entropy cost increases enormously when distances become really small. [Brillouin 13]

5. Further Applications

In addition to the application of information theory in determining the efficiency of a measurement, there are other conclusions which the theory of scientific information theory offers and which are stated without further qualification. [Mackay 33,34]

a) The various uncertainity relations of physics appear basically as axioms expressing the quantal nature of communicable information, consequent to the use of logical forms.

b) An experiment is not giving full information unless the metron-content of the <u>observation</u> (reading of a pointer) exceeds that of the <u>measurement</u> (characterized by apparatus, technique, and a priori structurization).

c) Performance of an experiment results fundamentally in the collection,

and allocation to the various logons, of the metron-flow arising from the impact of data on the apparatus plus observer.

d) In the statistical matching of one part of an experiment to another if a weak link in a sequence is known to yield only a certain metron content i_0, it is possible to estimate the time and/or space which it is worth-while to devote to each of the remaining links, and to gain in <u>overall-metron-content</u> per unit of space-time by designing these links so as to barter accuracy for speed or compactness.

e) If the total metrical information provided by a given technique is not usefully employed and is greater than the logon content, then to increase the selective information content it is more profitable to increase the logon content than the metron content.

f) In experiments to determine a constant, efforts should be directed toward "logon-compression" -- reducing the frequency response of the apparatus, with respect to time and space. In short, best results are obtained by acting consistently with one's belief that the constant will not alter with time or position, so that one logon will be sufficient.

g) In a sequence of operations, the logon-capacity of each should be adjusted so that the metron-content does not greatly exceed the value which it has in the stage with the narrowest bandwidth. This will enable each subsidiary operation to occupy the minimum space and time, so giving a higher overall metron-capacity, and making possible more repetitions of the experiment in a given space-time tract.

h) With a given input of energy, there is almost always an improvement in resolving power (structural detail) when intelligent steps are taken to sacrifice metrical information.

i) An increase in the metron-content of individual logons can be bought at the expense of logon-capacity, but the limit is set by the total metron content, which depends on the expanse of coordinate tract devoted to the experiment.

6. Parallelism between Communication Theory and Scientific Information Theory

Having now considered in a general way the applicability of information theory to scientific measurement and procedure, it will be appropriate to suggest a parallelism between features of a communication system between individuals and an information linkage in scientific measurement. Admittedly there are differences; the primary one being, as pointed out previously, that the goal of the communication system is <u>replication</u> and the goal of the scientific information linkage is <u>formulation</u>. In both instances a productive analysis necessitates including the source of information and the receiver of information in the system. In both, the a priori and a posteriori probabilities are factors to be considered. In both, the possibility of characterizing as many aspects of the system components as possible in mathematical terms augments the techniques whereby input, output and efficiency may be studied. On the following page, items in the left hand column are applicable to a general communication system. Analogous features of a scientific information system are listed in the right hand column.

COMMUNICATION SYSTEM	SCIENTIFIC INFORMATION SYSTEM
information source (an individual)	space-time tract of experimental interest (extra-observer)
code ensemble; probability constraints	"laws" of "nature"
message (selection from ensemble)	selection of experimental approach, devices, technique
fixed constraints (signal code organization)	logon capacity
transmitter operation (modulation and production of transmission signal)	transducer action of "immediately" influenced measurement device or component
channel	intermediate instrumentation and medium
channel capacity	metron capacity
noise - distortion (functional relation between transmitted and received signal)	systematic errors
noise - random	random errors
receiver operation (message reconstruction)	indicating device (classical observation level)
recipient (an individual)	observer
replication (selection from a priori ensemble; noise reduction; logical operations)	formulation (estimation of errors; compatibility with a priori laws of science; logical operations)
verification (repetition of transmission; alternate channel checks)	verification (comparison with previous results; alternate methods of experimental investigation)
compromise (band width, noise level)	compromise (metron content, logon content)

7. Summarization of Aims of Information Theory

Having examined and compared two applications of Information Theory, we may summarize its aims as follows:

a) to isolate from their particular contexts those abstract features of representations which can remain invarient under reformulation.

b) to treat quantitatively the abstract features of processes by which representations are made.

c) to give quantitative meanings to the several senses in which the notion of amount of information can be used.

With regard to scientific information theory, the realization of these aims embodies the consideration of all those factors which contribute to the formulation of the representation by the investigator, i.e., apparatus, scales of measurement, dimensions of measurement, the coupling of various components of the information system, extrapolation from one space-time scale of observation to another, errors, and the operations of the investigator pertinent to formulation of models, of scientific description, and the constraints of nature. Richards, speaking on the subject of language, has in a general way expressed the need to consider all the components of a system as far as possible.

> ..... The very instruments we use, if we try to say something which is not trivial about any aspect of language, embody in themselves the problems we hope to use them to explore..... All studies suffer from, and thrive through, this: that the properties of the instruments or apparatus employed enter into, contribute to, belong with, and confine the scope of the investigation..... I conjecture and I speak very humbly here -- that mathematics may have been the earliest study forced to ask itself about its own intellectual viewpoint, and the influence of its symbolism on its scope. This may suggest that the more abstract the properties of the instruments, the easier it may be to take account of their presence and not overlook them..... [Richards 43]

We have noted some of the applications and considerations of information theory; it has been emphasized that the potential elegance of the latter is rooted in sharpened definitions of the basic features of communication channels, which definitions are essential to a mathematical description of those features. However, the prospect of a more precise method of investigation should not cause one to overlook inherent limitations in the method of attack. In this regard the remarks of Fano (18) may well be heeded.

> One should also avoid confusing a physical system with the mathematical model which is used to represent it. The same physical system may be represented by different theoretical models, depending upon the problem under consideration. For instance, a computing machine may well be considered as a communication channel when certain aspects of its behavior are of interest, or as a perfectly determinate transducer when other aspects are the relevant ones. The fact is that we can never represent completely any physical system by means of a mathematical model because we cannot conceive of a model sufficiently complex; and even if we could conceive of it, it would be valueless to us because we could not analyze it.

CONCLUSION

The concept of entropy was considered in thermodynamics and in statistical mechanics as an aid to understanding the relationship of entropy and an amount of information. It was noted that the entropy of statistical mechanics and thermodynamic entropy were not identical in nature if the absolute validity of the second principle is assumed.

A definition of the amount of information received in a message was given:

$$I = \log \frac{Pea}{Peb}$$

It was demonstrated that this formula could be applied to events or to messages in a discrete communication system with or without noise. Different interpretations of "information" were brought forth, and it was seen that the more comprehensive analysis of Mackay resolved ambiguities. This analysis measures the amount of information in a message received according to its statistical rarity, and designates the result as the amount of selective information.

We then considered several problems in which "information" behaved, analytically speaking, as the negative of entropy. Despite the fact that this result appeared to agree with rough intuitive ideas in which entropy is deemed to be a type of "missing information," it was pointed out that "information" could be <u>identified</u> with the negative of <u>physical</u> entropy only in properly qualified systems. However, "selective information" could be <u>identified</u> with the non-thermodynamical form of entropy or the entropy of a set of probability distributions.

Information theory in its most general form embodied methods which were not necessarily limited to treatment of communication between individuals. If the communication application of information theory by its quantization and definition, in a manner susceptible to mathematical analysis, offered more elegant and productive methods of solving "communication" problems, could not similar methods be applied to other fields of endeavor which involve a transfer of information? Since one of the most important of the latter is the field of "Scientific Information" as it arises from experimentation on physical systems, it was deemed advantageous to recount various viewpoints on the problem of measurement in general. Although differences in these were apparent, it was proposed that, most generally, scientific measurement and formulation deal with an observer extracting information from a space-time tract by interaction and analysis.

To deal effectively with scientific measurement, information theory defines underlying phenomena in terms suitable for analytical and statistical treatment. The means whereby Scientific information theory "quantizes" scientific information was described in detail and was seen to be based upon the concepts of structural information and metrical information. These concepts in terms of logon and metron content make possible the assessment of the total information content for a given apparatus and technique, practical conclusions as to which factor (metron or logon) should be emphasized to gain specific results, and appreciation of the entropy cost for accuracy. The application of information theory to scientific measurement demonstrates that there are definitely aspects therein which parallel the communication problem --

[illegible]

[illegible]

[illegible]

[illegible]

i.e., the a priori and a posteriori representation; that in a measurement the various components should be compatible in discrimination and statistical nature for optimal efficiency in the same way that components of a communication system must be matched in their statistical features, and compatible in their characteristics for the maximum transfer of information within a given system.

There are two aspects of information theory which have been purposely omitted and yet which may be confounded with what has been set forth in this paper. In speaking of scientific information theory, no reference was made to <u>physical reality</u>; regarding "information," no reference was made to the <u>utility</u> for an individual <u>of information</u> received. Neither of these aspects are amenable to the techniques proposed in this paper.

BIBLIOGRAPHY

1. Baker, H. D., E. A. Ryder, and N. H. Baker — Temperature Measurement in Engineering. New York: John Wiley & Sons, 1953

2. Barnard, G. D. — "The Theory of Information," Journal of the Royal Statistical Society, Series B, XIII No. 1, 1951, 46-64

3. Bell, D. A. — Information Theory and its Engineering Applications. London: Sir Isaac Pitman & Sons, 1953

4. Bergmann, Gustav — "The Logic of Quanta," American Journal of Physics, 15, 1947, 397-408

5. Bergmann, P. G. — Basic Theories of Physics. New York: Prentice-Hall, 1951

6. Bohm, David — Quantum Theory. New York: Prentice-Hall, 1951

7. Bridgman, P. W. — The Nature of THERMODYNAMICS. Cambridge, Massachusetts: Harvard University Press, 1943

8. Bridgman, P. W. — Reflections of a Physicist. New York: Philosophical Library, 1950

9. Bridgman, P. W. — The Nature of Some of our Physical Concepts. New York: Philosophical Library, 1952

10. Brillouin, L. — "Life, Thermodynamics, and Cybernetics," American Scientist, 37, 1959, 554-568

11. Brillouin, L. — "Maxwell's Demon Cannot Operate: Information and Entropy. I.," Journal of Applied Physics, 22, 1951, 334-337

12. Brillouin, L. "Physical Entropy and Information. II.," Journal of Applied Physics, 22, 1951, 338-343

13. Brillouin, L. "The Negentropy Principle of Information," Journal of Applied Physics, 24, 1953, 1152-1163

14. Corben, H. C., and P. Stehle. Classical Mechanics. New York: John Wiley & Sons, 1950

15. Degroot, S. R. Thermodynamics of Irreversible Processes. New York: Interscience Publishers, 1951

16. Dingle, Herbert. "On the Dimensions of Physical Magnitudes," Philosophic Magazine, Series VII, Vol. 33, 1942, 321-344

17. Eddington, A. S. The Nature of the Physical World. New York: The Macmillan Company, 1929

18. Fano, R. M. Communication Theory. Edited by Willis Jackson, New York: Academic Press, 1953, 484

19. Finck, J. L. THERMODYNAMICS from a Generalized Standpoint. Brooklyn: Flatbush Publications, 1951

20. Fisher, R. A. The Design of Experiments. Fourth Edition, New York: Hafner-Publishing Company, 1947

21. Gabor, D. "A Summary of Communication Theory," Communication Theory, Edited by Willis Jackson, New York: Academic Press, 1953, 1-23

22. Gibbs, J. W. THE COLLECTED WORKS of J. Willard Gibbs. Volume I, Thermodynamics, New York: Longmans, Green and Co., 1928

23. Goldman, Stanford. Information Theory. New York: Prentice-Hall, 1953

24. Good, I. J. Probability and the Weighing of Evidence. London: Charles Griffin & Company, 1950

25. Heisenberg, Werner — Philosophic Problems of Nuclear Science. New York: Pantheon, 1952

26. Hercus, E. O. — Elements of Thermodynamics and Statistical Mechanics. Melbourne: University Press, 1950

27. Jordan, P. — "On the Process of Measurement in Quantum Mechanics," Philosophy of Science, 16, 1949, 269-278

28. Kantor, J. R. — The Logic of Modern Science. Bloomington, Indiana; Evanston, Illinois: The Principia Press, 1953

29. Kennard, E. H. — Kinetic Theory of Gases. New York: McGraw-Hill, 1938

30. Klein, J. F. — Physical Significance of Entropy. New York: D. Van Nostrand Company, 1910

31. Lindsay, R. B. — Introduction to Physical Statistics. New York: John Wiley & Sons, 1941

32. Lindsay, R. B. — Concepts and Methods of Theoretical Physics. New York: D. Van Nostrand Company, 1951

33. Mackay, D. M. — "Quantal Aspects of Scientific Information," Philosophical Magazine, 41, 1950, 289-311

34. Mackay, D. M. — "In Search of Basic Symbols," Cybernetics Transactions of the Eighth Conference, Edited by Heinz von Foerster, Caldwell, N. J.: Progress Associates, 1952

35. Margenau, Henry — The Nature of Physical Reality. New York: McGraw-Hill, 1950

36. Menzel, D. H. and David Layzer — "The Physical Principles of the Quantum Theory," Philosophy of Science, 16, 1949, 303-324

37. Muses, C. A. — An Evaluation of Relativity Theory. New York: S. Weiser, Inc., 1953

38. Oppenheimer, Robert — "The Age of Science," _Scientific American_, Sept. 1950, 20-23

39. Planck, Max — _Treatise on Thermodynamics_. Third Edition, London: Longmans, Green & Company, 1927

40. Raymond, R. C. — "Communication, Entropy, and Life," _American Scientist_, 38, 1950, 273-278

41. Raymond, R. C. — "Well Informed Heat Engine," _American Journal of Physics_, 19, 1951, 109-112

42. Reichenbach, Hans — _Philosophic Foundations of Quantum Mechanics_. Berkeley and Los Angeles: University of California Press, 1946

43. Richards, I. A. — "Communication between Men: The Meaning of Language," _Cybernetics Transactions of the Eighth Conference_, Edited by Heinz von Foerster, Caldwell, N. J.: Progress Associates, 1952

44. Schrodinger, Erwin — _Statistical Thermodynamics_. Cambridge: University Press, 1948

45. Sears, F. W. — _An Introduction to Thermodynamics, The Kinetic Theory of Gases, and Statistical Mechanics_, Second Edition. Cambridge, Massachusetts: Addison-Wesley Publishing Company, 1953

46. Shannon, C. E. — "A Mathematical Theory of Communication," _The Bell System Technical Journal_. 27, 1948, 379-423, 623-656

47. Smith, R. A. — _The Physical Principles of Thermodynamics_. London: Chapman & Hall, 1952

48. Stevens, — "On the Theory of Scales of Measurement," _Science_, 34, 1946

49. Tolman, R. C. — _The Principles of Statistical Mechanics_. London: Oxford University Press, 1938

50. von Bertalanffy, Ludwig — "The Theory of Open Systems in Physics and Biology," _Science_. 3, 1950, 23-29

51. Watanabe, M. S. — "A Study of Ergodicity and Redundancy Based on Intersymbol Correlation of Finite Range." U. S. Naval Post-Graduate School: 1954

52. Wiener, Norbert — _Cybernetics_. New York: John Wiley & Sons, 1948

53. Wilson, W. — "Dimensions of Physical Quantities," _Philosophical Magazine_, Seventh Series, Vol. 33, 1942, 26-33

54. Whittaker, E. T. — "Some Disputed Questions in the Philosophy of Physical Sciences," _Philosophical Magazine_, Seventh Series, Vol. 33, 1942, 353-366

55. Woodward, P. M. — _Probability and Information Theory, with Applications to Radar_. New York: McGraw-Hill, 1953

APPENDIX I

LIST OF DEFINITIONS

1. Representation -- a representation is any structure (pattern, picture, model), whether abstract or concrete, of which the features purpost to symbolize or correspond in some sense with those of some other structure.
2. Structural Information Content -- this quantity is defined as the number of distinguishable groups or clusters in a representation -- the number of definably independent respects in which it could vary -- its dimensionality or number of degrees of freedom or basal multiplicity.
3. Logon -- the unit of structural information, one logon, is that which enables one such new distinguishable group to be defined for a representation.
4. Logon Content -- this is a convenient term for the structural information content or number of logons (number of independently variable features) in a representation (e.g., the number of independent coefficients required to specify a given wave form over a given period of time).
5. Metrical Information Content -- the definition of this term is: the number of (indistinguishable) logical elements in a given group or in the total pattern.
6. Metron -- the unit of metrical information, one metron, is defined as that which supplies one element for a pattern. Each

element may be considered to represent one unit of evidence. Thus the amount of metrical information in a pattern measures the *weight of evidence* to which it is equivalent. Metrical information gives a pattern its weight or density -- the "stuff" out of which the "structure" is formed.

[Mackay 31]